Thermoplasma acidophilum citrate synthase: cloning and sequencing of the gene

Sutherland, Katharine J.

Award date:
1990

Awarding institution:
University of Bath

Thermoplasma acidophilum CITRATE SYNTHASE: CLONING AND SEQUENCING OF THE GENE

submitted by Katharine J. Sutherland
for the degree of PhD
of the University of Bath 1990

Kate Sutherland.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To dad and mum and Helen

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr. Michael Danson, Dr. David Hough and Dr. Paul Towner, for their help and encouragement throughout the project. Thanks are also due to Dr. Len Hall and Dr. Derrick Bowen in Bristol for their assistance during the early stages of the project. In addition, I would like to acknowledge that the sequence alignments (Chapter 5) were carried out by Dr. Christina Henneke and the secondary structure predictions (Chapter 5) were carried out by Dr. Gary Taylor.

Lastly mention should go to the people of the "middle" and "end" labs for all their help and support during my stay in Bath.
Thank you!

ABBREVIATIONS

Most of the abbreviations used in this thesis are those recommended in the "Policy of the Biochemical Journal and Instructions to Authors", (1990) Biochem. J. 265, 1-21.

Ap^r, ampicillin resistance
BSA, bovine serum albumin
CoA, coenzyme A
DTNB, 5,5'-dithio-bis-(2-nitrobenzoic acid)
DTT, dithiothreitol
EDTA, (disodium) ethylenediaminetetraacetate
IPTG, isopropyl-ß-D-thiogalactoside
Kb, 1000 base-pairs
K_i, inhibition constant
K_m, Michaelis constant
LMP, low melting point
M_r, relative molecular mass
NAD, nicotinamide-adenine dinucleotide
OD, optical density
ORF, open reading frame
PAGE, polyacrylamide-gel electrophoresis
PEG, polyethylene glycol
rRNA, ribosomal RNA
RT, room temperature
SDS, sodium dodecyl sulphate
SSC, 0.15 M sodium chloride and 15 mM sodium citrate
TBE, 90 mM Tris base, 90 mM boric acid and 2 mM EDTA
TE, 10 mM Tris-HCl and 1 mM EDTA
TEA, triethylamine
TEMED, N,N,N',N' tetramethylethylene diamine
T_m, temperature of hybridisation
X-gal, 5-bromo-4-chloro-3-indolyl-ß-D-galactoside
3D, 3-dimensional

Enzymes

Citrate synthase or CS, citrate synthase (EC 4.1.3.7)

EF-Tu, elongation factor Tu

GAPDH, glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.-)

GDH, glucose dehydrogenase (EC 1.1.1.47)

GS, glutamine synthetase (EC 6.3.1.2)

MDH, malate dehydrogenase (EC 1.1.1.37)

RNAse, ribonuclease A (EC 3.1.27.5)

SUMMARY

In order to extend our knowledge of archaebacterial citrate synthases, the gene for this enzyme was cloned from the thermoacidophilic archaebacterium, Thermoplasma acidophilum, into Escherichia coli. Two oligonucleotides, a short redundant 17-mer and a long non-redundant 48-mer, both based on the N-terminal amino acid sequence of Tp.acidophilum citrate synthase, were used to probe Southerns of Tp.acidophilum DNA. Only the CS-48-mer proved specific for the citrate synthase gene.

An initial construct, pTaCS1, isolated as a result of cloning a 1.6-kb fragment of Tp.acidophilum DNA, carried only a truncated 5′ portion of the citrate synthase gene. An oligonucleotide (CS-16-mer), based on sequence information from pTaCS1, was used to probe for the remaining 3′ portion of the gene and the construct, pTaCS2, containing this desired region was isolated. A third construct, pTaCS19, carrying the complete Tp.acidophilum citrate synthase gene, was assembled by combining the 5′ and 3′ portions (from pTaCS1 and pTaCS2) at a common SacI site.

The Tp.acidophilum citrate synthase gene and its immediate 5′ and 3′ flanking regions were sequenced. The derived amino acid sequence was aligned with other citrate synthases from eubacterial and eukaryotic organisms. The percentage identities between the archaebacterial and non-archaebacterial sequences was found to be less than 30%. Nevertheless, the majority of residues implicated in the catalytic action of the enzyme appear to have been conserved across all three kingdoms.

The flanking regions of the Tp.acidophilum citrate synthase gene were compared with archaebacterial consensus sequences; putative regions involved in initiation and termination of transcription and translation are proposed.

The Tp.acidophilum gene was expressed in E.coli to produce catalytically active citrate synthase. The

enzyme was purified using a two-step purification procedure: a heat-treatment at 65°C followed by a chromatofocussing step.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

1.1. The citric acid cycle

The formation of citrate from oxaloacetate and pyruvate was first demonstrated by Krebs and Johnson (1937). This was the final discovery in the elucidation of a series of reactions which we now call the citric acid cycle (or alternatively the tricarboxylic acid or the Krebs' cycle). With the determination of the roles of acetyl-CoA and succinyl-CoA by Lynen & Reichert (1951) the scheme for the citric acid cycle, as we know it today, was complete. The main steps of the cycle are summarised in Fig. 1.1.

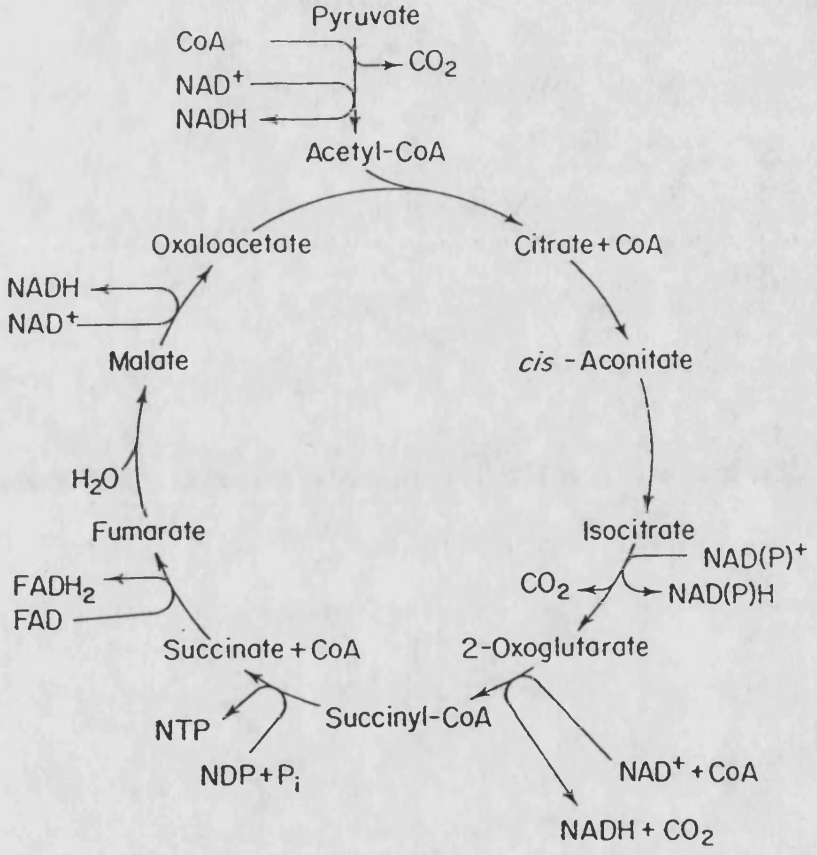
The citric acid cycle acts as the pathway for the final oxidation of all major nutrients in most living organisms [Lowenstein (1967)] and has two vital functions: (i) the generation of energy via the production of $NAD(P)H_2$ and (ii) the formation of intermediates for biosynthesis [Krebs *et al.* (1952); Roberts *et al.* (1953)]. Diverse organisms have different requirements for the citric acid cycle as a result of each species having its own distinct energy and biosynthetic needs. Consequently, citric acid cycle enzymes (isolated from different sources) display great diversity in terms of activity, structure and regulatory sensitivities [reviewed by Weitzman (1981); Danson (1988)]. The diversity of citrate synthase, the first enzyme of the citric acid cycle, is discussed below.

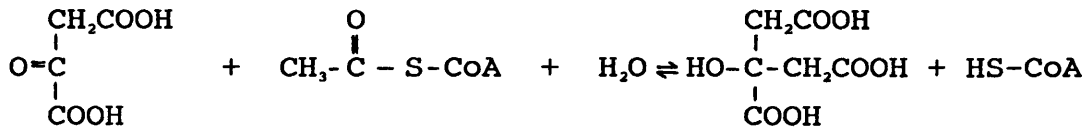
1.2. Citrate synthase

1.2.1. The role of citrate synthase

Citrate synthase, first described by Stern *et al.* (1950), catalyses the condensation of oxaloacetate and acetyl-CoA to form citrate, thus facilitating the entry of carbon into the citric acid cycle. The reaction catalysed by citrate synthase is:

Figure 1.1: The citric acid cycle





The majority of citrate synthases have a si-stereospecificity: the acetyl group of acetyl-CoA being added to the si-face of the keto moiety of oxaloacetate, and the methyl group of the acetyl-CoA undergoing an inversion [Wiegand & Remington (1986) and references therein]. The overall reaction can be dissected into three partial reactions, namely (a) enolisation: base abstraction of a methyl proton from acetyl-CoA; (b) condensation: nucleophilic attack of the carbanion of acetyl-CoA on the carbonyl group of oxaloacetate to produce citryl-CoA; and (c) hydrolysis: hydrolysis of citryl-CoA to produce citrate and CoA [Wiegand & Remington (1986) and references therein].

Citrate synthase, the only enzyme in the citric acid cycle involved in the formation of a carbon-to-carbon bond, catalyses the production of a 6-carbon compound (citrate) with new metabolic potential. This step, crucial to the cycle's "dual function" in biosynthesis and energy production, makes citrate synthase a potentially important point of control of the citric acid cycle. The regulation of citrate synthases is discussed in Sections 1.2.3 and 1.2.4.

1.2.2. The diversity of citrate synthase

Citrate synthase has been studied from organisms that represent all three primary kingdoms (the eukaryotes, the eubacteria and the archaebacteria: see Section 1.3). The enzyme has been found to possess diversity with regard to structure, and to kinetic and regulatory behaviour [Srere (1972); Weitzman & Danson (1976); Danson (1988)]. This diversity seems to be correlated with the taxonomic status of the source organism [Weitzman & Jones (1968); Weitzman & Danson (1976); Danson (1988)].

Two oligomeric forms of citrate synthase have been identified: a dimeric form found in eukaryotes, Gram-positive eubacteria and archaebacteria, and a hexameric form found in Gram-negative eubacteria [see Weitzman (1981) and references therein; Danson et al. (1985)]. Both types are made up of identical subunits with M_r values of approximately 50 000, the hexameric form appearing to behave functionally as a trimer of the basic dimer [Else et al. (1988)].

1.2.3. Eukaryotic and eubacterial citrate synthases

Citrate synthases of eubacteria and eukaryotes have been found to differ in their regulatory sensitivities to various nucleotides. The "large" hexameric citrate synthase of Gram-negative eubacteria is allosterically inhibited by NADH, an effect first observed for the Escherichia coli enzyme by Weitzman (1966) and since reported for a large number of other Gram-negative eubacterial citrate synthases [Weitzman & Jones (1968); Weitzman (1981) and references therein]. The "small" dimeric citrate synthase of Gram-positive eubacteria and eukaryotes, on the other hand, is not allosterically inhibited by NADH, but instead appears to be isosterically inhibited by ATP [Hathaway & Atkinson (1965); Weitzman & Danson (1976) and references therein]. NADH as an end-product, and ATP as an ultimate end-product of the citric acid cycle, can both be envisaged as acting as feedback regulators of energy production. It should be noted, however, that the in vivo significance of ATP inhibition is questionable since Mg^{2+} ions, at physiological concentrations, have been observed to cancel the ATP effect [Weitzman & Hewson (1973)].

An additional control, by AMP, has been observed for some Gram-negative citrate synthases; the NADH-inhibited enzyme from obligate aerobes is reactivated by AMP, whereas the equivalent form from facultative anaerobes is insensitive to this nucleotide [Weitzman & Jones (1968)]. The need for this additional control in aerobes can be

rationalised by the fact that these organisms depend on the citric acid cycle for energy production; whereas, the anaerobes can generate energy by fermentation.

A biosynthetic intermediate of the citric acid cycle, 2-oxoglutarate, also acts as a regulator of some citrate synthases [Wright et al. (1967)]. This 2-oxoglutarate inhibition is limited to organisms which lack 2-oxoglutarate dehydrogenase and so produce 2-oxoglutarate as the end-product of the citrate synthase arm of the cycle [Weitzman & Dunmore (1969)]. The effect of 2-oxoglutarate can be envisaged as a typical example of feedback inhibition.

Table 1.1 summarises the properties of citrate synthases of eubacteria and eukaryotes.

1.2.4. Archaebacterial citrate synthases

Citrate synthase has been found in all three archaebacterial phenotypes: halophiles, methanogens and thermophiles [Danson et al. (1985)] (see Section 1.3 for discussion on archaebacteria). As mentioned above, the archaebacterial citrate synthases, like those of the Gram-positive eubacteria and the eukaryotes, are of the "small" type [Danson et al. (1985)]. Their regulation also appears to be similar to that of the Gram-positive and eukaryotic citrate synthases; the enzyme from both thermophilic and halophilic archaebacteria being sensitive to ATP (the halophilic enzyme only weakly so) and neither form showing allosteric inhibition by NADH or 2-oxoglutarate [Grossebüter & Görisch (1985); Danson et al. (1985)]. The regulatory behaviour of the methanogenic citrate synthase is not easy to determine since the enzyme is present in only low activity [Danson et al. (1985)].

Table 1.1 summarises the properties of citrate synthases of archaebacteria.

Table 1.1: Properties of citrate synthases from eubacteria, eukaryotes and archaebacteria

	Eubacteria		Eukaryotes	Archaebacteria
	Gram -ve	Gram +ve		
Native M_r	~290 000	~90 000	~100 000	80-110 000
Oligomeric nature	6n	2n	2n	2n
Allosteric inhibitors	NADH	-	-	-
K_iATP	>5 mM	~1 mM	<1 mM	1-4 mM

Data are from Danson (1988).

1.2.5. Current knowledge of citrate synthase primary and tertiary structures

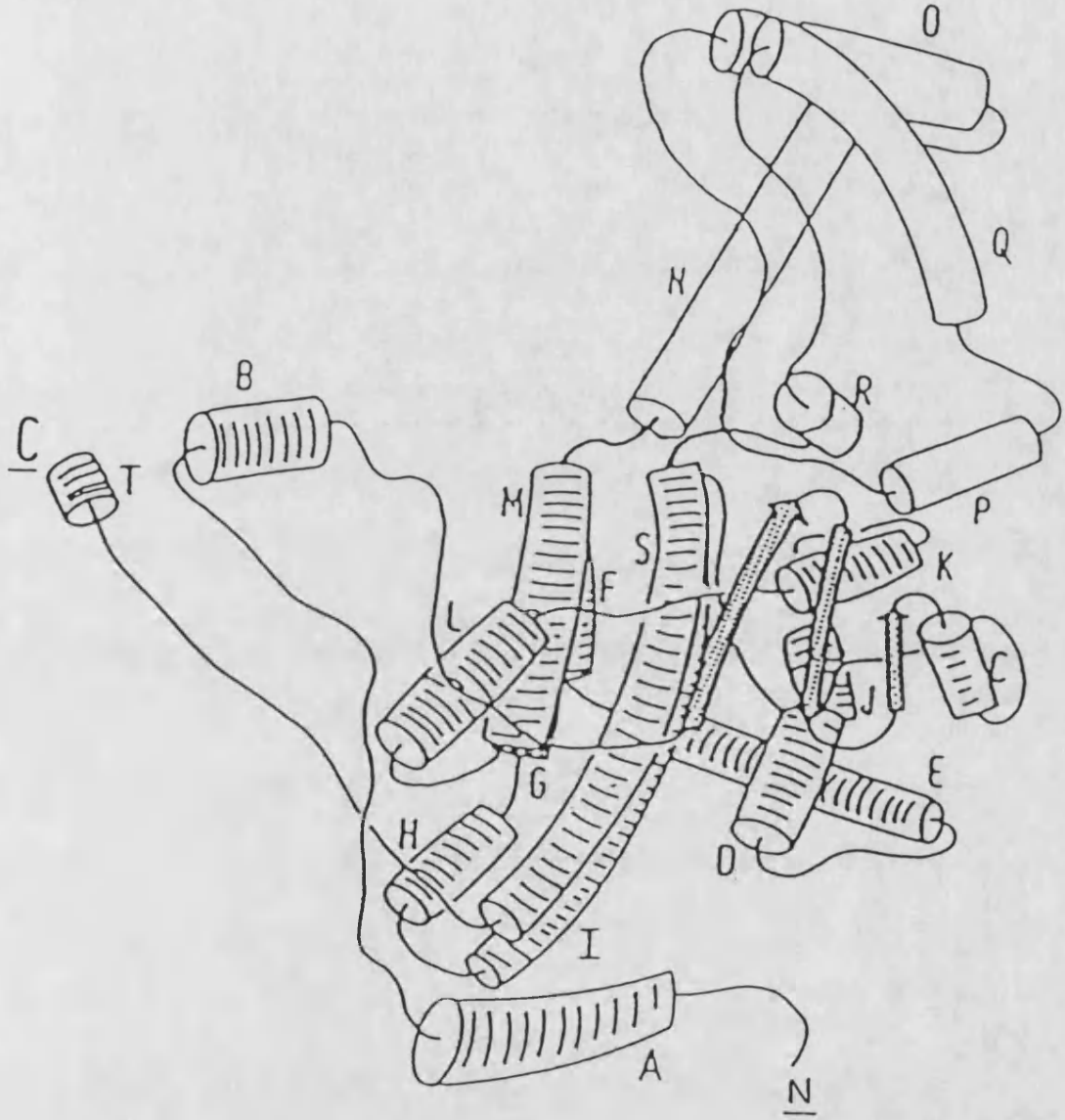
The interest in citrate synthase as a model enzyme to study structure, function and regulation has stimulated a vast amount of research. Currently, DNA sequences for eight citrate synthase genes and a tertiary structure for a eukaryotic citrate synthase are available.

The primary amino acid sequences of citrate synthase (determined by amino acid sequencing or from the DNA sequence) are available for the following eukaryotic citrate synthases: pig heart [Bloxham et al. (1981)], pig kidney [Evans et al. (1988)], Arabidopsis thaliana [Unger et al. (1989)], Saccharomyces cerevisiae mitochondria [Suissa et al. (1984)] and S.cerevisiae glyoxysome [Rosenkrantz et al. (1986)]; and for the following eubacterial citrate synthases: E.coli [Bhayana & Duckworth (1984); Ner et al. (1983)], Rickettsia prowazekii [Wood et al. (1987)], Acinetobacter anitratum [Donald & Duckworth (1987)] and Pseudomonas aeruginosa [Donald et al. (1989)].

A high-resolution X-ray crystallographic structure is available for the pig heart citrate synthase [Remington et al. (1982); Wiegand et al. (1984); Wiegand & Remington (1986)]. Each subunit consists of twenty α-helices. The residual segments form extended irregular structures, except for a very small ß-sheet of thirteen residues (Fig. 1.2).

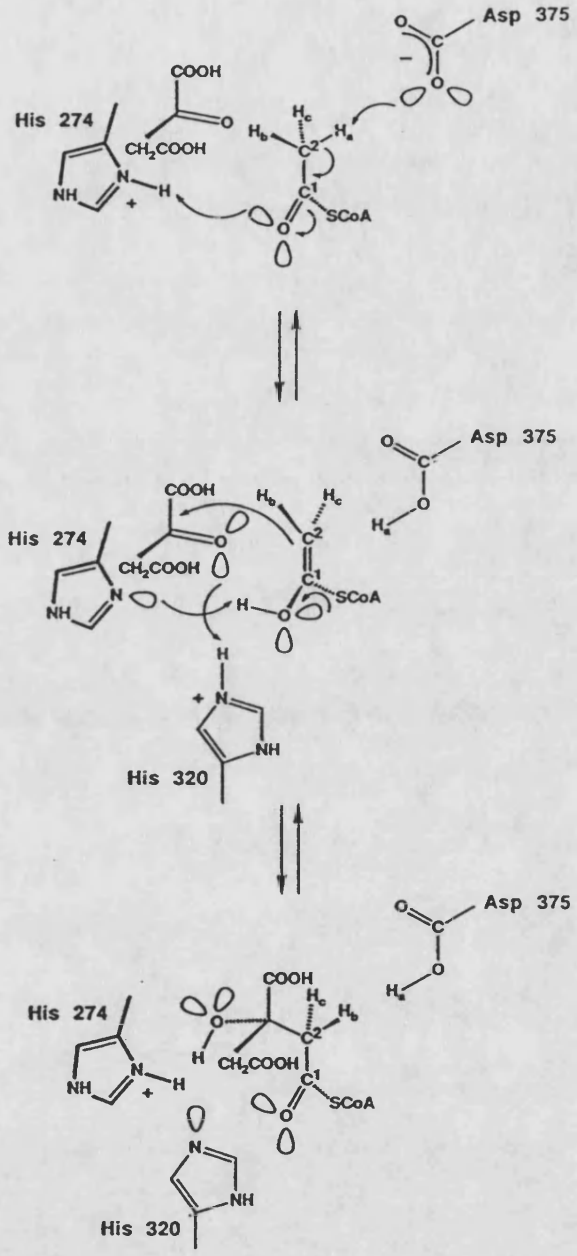
Three residues His 274, His 320 and Asp 375 are believed to be involved in the catalytic mechanism of citrate synthase [Remington et al. (1982)], and a scheme for the involvement of these residues during the condensation reaction of the enzyme has recently been proposed by Karpusas et al. (1990) (Fig. 1.3). The catalytic importance of His 274 and Asp 375 of the pig citrate synthase [Evans et al. (1989)], and of Asp 362 of the E.coli citrate synthase (believed to correspond to Asp 375 of the pig enzyme) [Handford et al. (1988)], has

Figure 1.2: Secondary structure of the pig citrate synthase



α-helices are labelled (A) to (T).
Reproduced from Remington et al. (1982).

Figure 1.3: Mechanism proposed for the condensation reaction of pig citrate synthase



Reproduced from Karpusas et al. (1990). The mechanism is proposed to proceed via simultaneous general acid-base catalysis (top) to form a neutral enol intermediate (centre) which then attacks the carbon of the oxaloacetate (bottom).

been demonstrated by site-directed mutagenesis and kinetic analysis.

Henneke et al. (1989) have developed a multiple alignment program for citrate synthase sequences, which uses a McLachlan simscore matrix and gives penalties for breaks and insertions within regions of secondary structure (these latter regions were deduced from the known tertiary structure of the pig enzyme). On a range of criteria (including conservation of residues believed to be involved in the active site of the pig enzyme), this matrix and the associated penalties were found to give the most realistic alignments for the eubacterial and eukaryotic citrate synthase sequences then available [Henneke et al. (1989)].

1.3. The archaebacteria

1.3.1. The concept of archaebacteria

Our views of cellular evolution were totally transformed in the late 1970's by the work of Woese and Fox (1977). Initially using partial ribosomal RNA (rRNA) sequences obtained by oligonucleotide cataloguing, and more recently using complete rRNA sequences [Lane et al. (1985)], Woese's group showed that life is not fundamentally of two types, prokaryotes and eukaryotes, but instead comprises three types, the eubacteria, the archaebacteria and the eukaryotes* (Fig. 1.4).

At this point it may be worth considering some of the reasons why 16S rRNA (18S in eukaryotes) was chosen as a phylogenetic marker [reviewed by Woese (1985, 1987)]. A phylogenetic marker must meet two essential criteria: universal distribution, and constancy in function. (The latter criterion ensures that any changes in sequence are selectively neutral.) The ribosome appears to have a constant function in all organisms, and

* see footnote on page 12

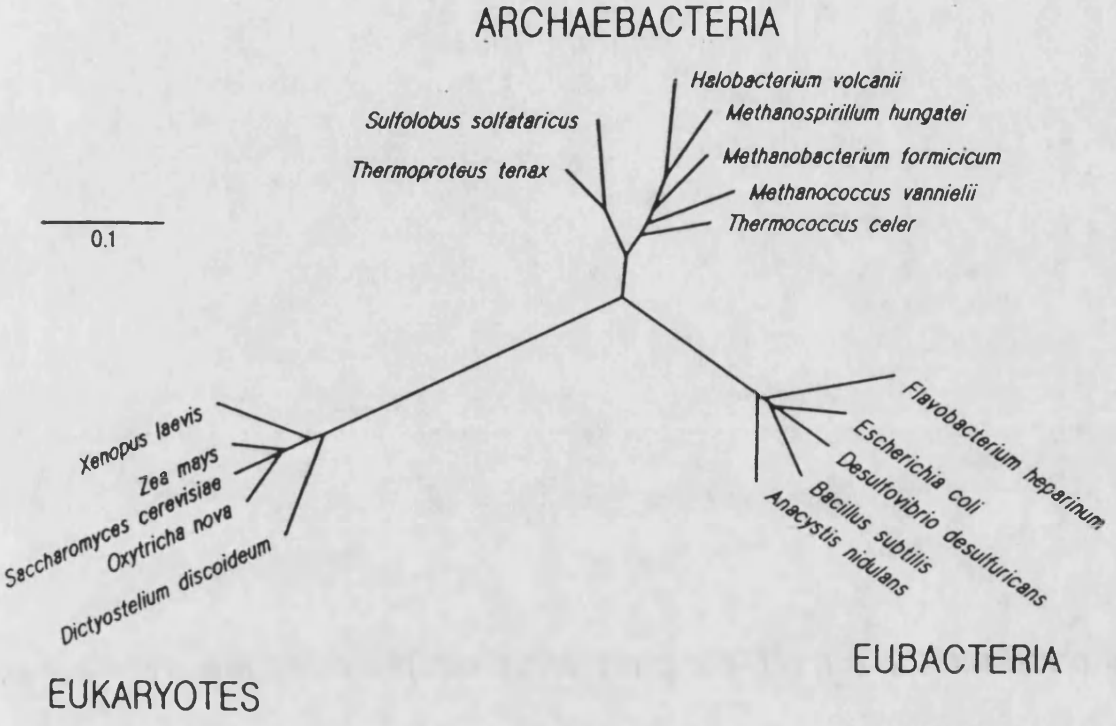
Figure 1.4: Woese's phylogenetic trees based on rRNA sequence comparisons

(a) Unrooted [reproduced from Woese & Olsen (1986)].

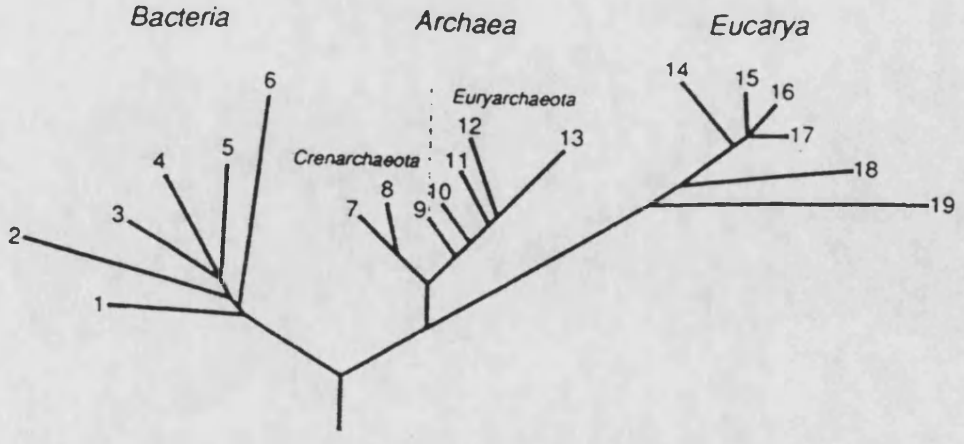
(b) Rooted [reproduced from Woese *et al.* (1990)]:

1. the Thermotogales; 2. the flavobacteria; 3. the cyanobacteria; 4. the purple bacteria; 5. the Gram-positive bacteria; 6. the green nonsulphur bacteria; 7. *Pyrodictium*; 8. *Thermoproteus*; 9. the Thermococcales; 10. the Methanococcales; 11. the Methanobacteriales; 12. the Methanomicrobiales; 13. the extreme halophiles; 14. the animals; 15. the ciliates; 16. the green plants; 17. the fungi; 18. the flagellates; and 19. the microsporidia.

(a)


ARCHAEBACTERIA
Halobacterium volcanii
Methanospirillum hungatei
Methanobacterium formicicum
Methanococcus vannielii
Thermococcus celer
Sulfolobus solfataricus
Thermoproteus tenax
0.1
Xenopus laevis
Zea mays
Saccharomyces cerevisiae
Oxytricha nova
Dictyostelium discoideum
Flavobacterium heparinum
Escherichia coli
Desulfovibrio desulfuricans
Bacillus subtilis
Anacystis nidulans
EUKARYOTES
EUBACTERIA

(b)


Bacteria
Archaea
Eucarya
Euryarchaeota
Crenarchaeota
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

thus can be seen to conform to both the above criteria. In addition, 16S rRNA, being a large molecule, contains regions of sequence that change either at a slow or a fast rate; therefore, whereas the former regions can be used to measure distant phylogenetic relationships, the latter can be used to measure closer relationships. Finally, rRNA can be sequenced directly and rapidly using reverse transcriptase [Lane et al. (1985)]. For these reasons then, the 16S/18S rRNA molecule is considered an excellent phylogenetic marker and currently remains the most useful.

1.3.2. The archaebacterial tree

The archaebacterial kingdom comprises three major groups: the methanogens, the sulphur-dependent thermophiles and the extreme halophiles* [Woese et al. (1978)]. The methanogens are strict anaerobes that produce methane from hydrogen and carbon dioxide, or other simple C_1 compounds [reviewed by Jones et al. (1987)]. The sulphur-dependent thermophiles grow at high temperatures (usually anaerobically) and utilise sulphur as an electron acceptor [reviewed by Stetter & Zillig (1985)]. The extreme halophiles are noted for their growth in high salt conditions, and their high intracellular salt concentrations of > 3.5 M [reviewed by Kushner (1985)].

On the basis of rRNA sequence comparisons, the archaebacteria fall into two major branches: (i) the sulphur-dependent thermophile branch; and (ii) the methanogen-halophile branch. The latter also contains

* New names - Bacteria (eubacteria), Archaea (archaebacteria), Eucarya (eukaryotes), Euryarchaeto (halophiles/methanogens) and Crenarchaeaota (sulphur-dependent thermophiles) - have recently been proposed by Woese et al. (1990).

two types of thermoacidophiles: Thermococcus celer and Thermoplasma acidophilum (Fig. 1.5) [Woese & Olsen (1986)]. Tp.acidophilum will be discussed further in Section 1.4.

The main characteristics of representatives for the three archaebacterial groups are summarised in Table 1.2.

1.3.3. Evidence supporting the uniqueness of archaebacteria

Although the phylogenetic distinctiveness of the archaebacteria was first recognised by Woese and colleagues through the use of 16S rRNA as a phylogenetic marker, their status as a third primary kingdom has been amply demonstrated by studies on their physiology and biochemistry, and by the use of other molecules as phylogenetic markers. (The use of proteins as phylogenetic markers will be discussed in Chapter 5.) In addition to possessing various unique characteristics, the archaebacteria share a number of features found either in the eukaryotes or in the eubacteria. These are summarised in Table 1.3. For reviews see Kandler (1982); Jones et al. (1987); Brown et al. (1989).

1.3.4. An alternative to Woese's archaebacterial tree

Woese's archaebacterial tree has not gone without criticism and perhaps the most controversial and intractable critic, James A. Lake, has proposed his own phylogenetic tree as "an alternative to archaebacterial dogma" (Fig. 1.6) [Lake (1987); Lake (1989)]. Based on the same 16S rRNA sequence data as that used by Woese, but using slightly different regions within this molecule, and employing a different tree reconstruction algorithm, Lake has proposed that the halobacteria, methanogens and eubacteria form one urkingdom and the sulphur-dependent thermophiles (termed the eocytes) and the eukaryotes form another. Lake claims that Woese's

Figure 1.5: The archaebacterial tree

Reproduced from Klenk et al. (1986).

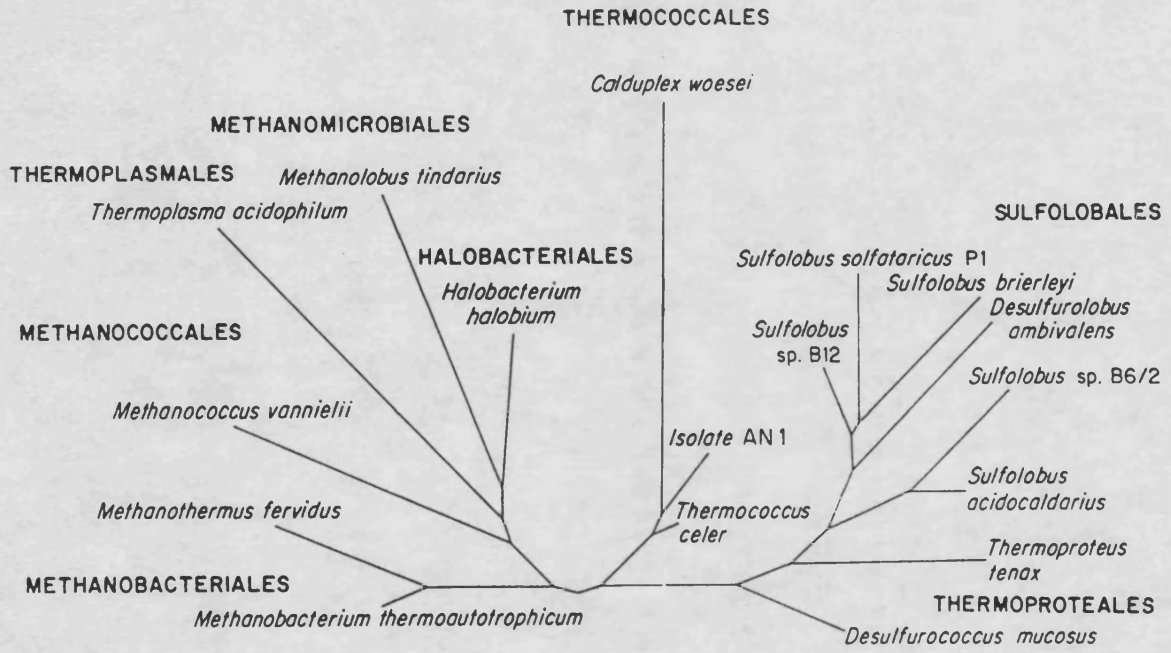


Table 1.2: Characteristics of some archaebacterial genera

Information is from Jones et al. (1987) and Hough & Danson (1989). The list is in not exhaustive.
ND: not determined.

Phenotype	Genera	Relation to O_2	Optimal NaCl(M)	G+C (mol%)	Temp opt(°C)	pH opt
Halophilic	Halobacterium	aerobic	2.0-4.5	60-68	35-50	6.5-7.2
	Halococcus	aerobic	3.5-4.5	65	30-37	7.2
	Natronobacterium	aerobic	3.0-3.5	63-65	37-45	7.2
	Natronococcus	aerobic	3.5-4.0	64	35-40	9.5
Thermo-philic	Thermoplasma	facultatively aerobic	-	38-46	59-60	1.0-2.0
	Sulfolobus	aerobic	-	30-40	70-90	1.5-5.0
	Archaeoglobus	anaerobic	-	46	83	5.5-7.5
	Thermococcus	anaerobic	-	51	92	5.8
	Thermoproteus	anaerobic	-	55	88	5.5
	Thermophilum	anaerobic	-	57	88	5.5
	Thermodiscus	anaerobic	-	53	87	5.5
	Desulfurococcus	anaerobic	-	51	85	6.0
	Pyrodictium	anaerobic	-	50-62	105	5.0-7.0
	Pyrococcus	anaerobic	-	ND	70-103	5.0-9.0
Methano-genic	Methanococcus	anaerobic	-	29-34	30-85	6.0-9.0
	Methanothrix	anaerobic	-	52	37-60	7.4-7.8
	Methanosarcina	anaerobic	-	39-42	35-50	6.0-7.0
	Methanomicrobium	anaerobic	-	45-49	40	6.1-7.0
	Methanogenium	anaerobic	-	47-61	20-60	6.2-7.5
	Methanobacterium	anaerobic	-	32-61	37-70	7.0-8.5
	Halomethanococcus	anaerobic	2.0-3.0	49	35	7.5
	Methanohalophilus	anaerobic	0.2-2.0	ND	30-45	9.0

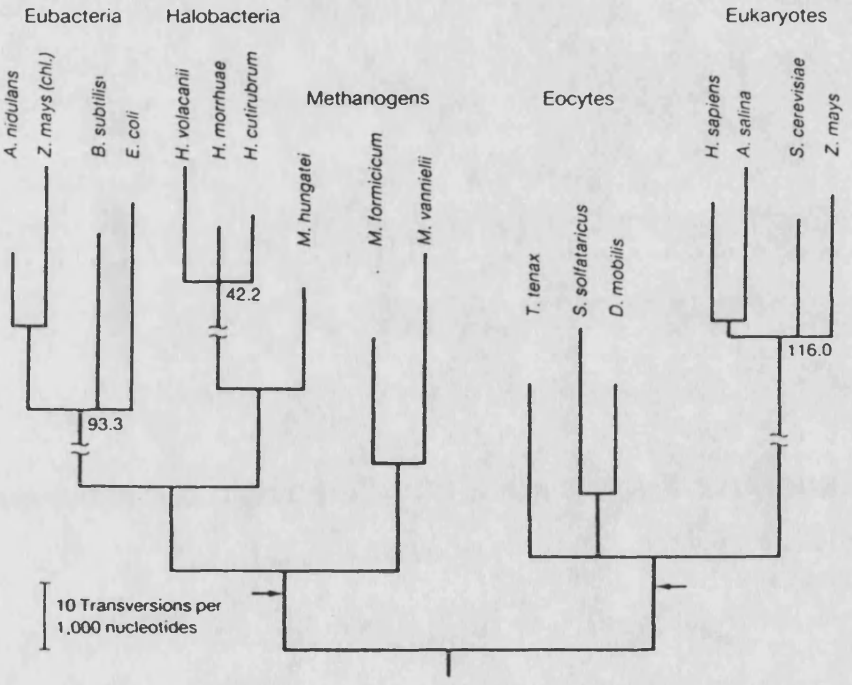
Table 1.3: Unique and shared (shaded) features of archaebacteria, eubacteria and eukaryotes

	ARCHAEBACTERIA	EUBACTERIA	EUKARYOTES
CELL SIZE (LINEAR DIMENSION)	ABOUT 1 MICROMETER	ABOUT 1 MICROMETER	ABOUT 10 MICROMETERS
CELLULAR ORGANELLES	ABSENT	ABSENT	PRESENT
NUCLEAR MEMBRANE	ABSENT	ABSENT	PRESENT
CELL WALL	VARIETY OF TYPES; NONE INCORPORATES MURAMIC ACID	VARIETY WITHIN ONE TYPE; ALL INCORPORATE MURAMIC ACID	NO CELL WALL IN ANIMAL CELLS; VARIETY OF TYPES IN OTHER PHYLA
MEMBRANE LIPIDS	ETHER-LINKED BRANCHED ALIPHATIC CHAINS	ESTER-LINKED STRAIGHT ALIPHATIC CHAINS	ESTER-LINKED STRAIGHT ALIPHATIC CHAINS
TRANSFER RNA'S:			
THYMINE IN "COMMON" ARM	ABSENT	PRESENT IN MOST TRANSFER RNA'S OF MOST SPECIES	PRESENT IN MOST TRANSFER RNA'S OF ALL SPECIES
DIHYDROURACIL	ABSENT IN ALL BUT ONE GENUS	PRESENT IN MOST TRANSFER RNA'S OF ALL SPECIES	PRESENT IN MOST TRANSFER RNA'S OF ALL SPECIES
AMINO ACID CARRIED BY INITIATOR TRANSFER RNA	METHIONINE	FORMYLMETHIONINE	METHIONINE
RIBOSOMES:			
SUBUNIT SIZES	30S, 50S	30S, 50S	40S, 60S
APPROXIMATE LENGTH OF 16S (18S) RNA	1,500 NUCLEOTIDES	1,500 NUCLEOTIDES	1,800 NUCLEOTIDES
APPROXIMATE LENGTH OF 23S (25-28S) RNA	2,900 NUCLEOTIDES	2,900 NUCLEOTIDES	3,500 NUCLEOTIDES OR MORE
TRANSLATION-ELONGATION FACTOR	REACTS WITH DIPHTHERIA TOXIN	DOES NOT REACT WITH DIPHTHERIA TOXIN	REACTS WITH DIPHTHERIA TOXIN
SENSITIVITY TO CHLORAMPHENICOL	INSENSITIVE	SENSITIVE	INSENSITIVE
SENSITIVITY TO ANISOMYCIN	SENSITIVE	INSENSITIVE	SENSITIVE
SENSITIVITY TO KANAMYCIN	INSENSITIVE	SENSITIVE	INSENSITIVE
MESSENGER-RNA BINDING SITE AUCACCUCC AT 3' END OF 16S (18S) RNA	PRESENT	PRESENT	ABSENT

Reproduced from Woese (1981).

Figure 1.6: Lake's Eocyte tree based on rRNA sequence comparisons

Reproduced from Lake (1988).



algorithms fail to account adequately for the "unequal rate effect"*, a phenomenon that results in clustering of unrelated organisms.

Woese has counteracted with a detailed critique of Lake's work [Olsen & Woese (1989)]. He claims that Lake's eocyte tree is an artifact due to "an (unspecified) systematic error" in the analysis.

Although the Lake-Woese controversy remains unsolved, in view of the overwhelming physiological and biochemical evidence in support of Woese's archaebacterial tree, the Woese view of phylogeny remains generally accepted.

1.3.5. The Universal Ancestor

The first reasonable attempt to describe the origins of a living cell were presented by Oparin in 1924 [reviewed by Woese (1987)]. This theory developed the notion that the earliest organisms were heterotrophic prokaryotes which later evolved photosynthetic, autotrophic and aerobic capabilities. Eukaryotic cells were believed to be more advanced in evolutionary terms, having formed from wall-less prokaryotes that gained the capacity for endocytosis [Margulis (1970)]. Such theories, however, were the result of speculation and it has only been since the advent of sequencing techniques that a more objective approach to determining the nature of a universal ancestor has been made possible.

In Woese's phylogenetic tree all three basic cell types are thought to have diverged from a common ancestor; consequently the eukaryotic cell is envisaged to be as ancient as that of the two prokaryotic cell types. However, the restrictive nature of archaebacterial environments is believed to have caused

* "The unequal rate effect" is caused by the rates of change of a nucleotide being different in adjacent branches of a phylogenetic tree.

the archaebacteria to evolve more slowly and, as a result, the archaebacterial phenotype is often postulated to be that which most closely resembles the universal ancestor. Moreover, since thermophilic species have a widespread distribution across the phylogenetic trees of both Woese and Lake, it seems unlikely that these organisms arose from a mesophilic ancestor. Consequently, both Woese and Lake have inferred that, among the organisms seen today, the archaebacterial sulphur-dependent thermophile is probably the phenotype that most resembles that of the ancestral organism [Achenbach-Richter et al. (1987); Oyaizu et al. (1987); Lake (1989)].

1.3.6. Archaebacteria and biotechnology

It is interesting to consider what adaptations are required for life to survive under extreme conditions (such as high and low temperatures, high and low pH, high pressure, and high salt) from both an intellectual point of view and a biotechnological one. The archaebacterial kingdom offers numerous "good" examples of organisms living under extreme conditions. For example, the maximum temperature of growth for a eubacterial species (i.e. Thermotoga maritima) is 90°C [Huber et al. (1986)], whereas archaebacteria (i.e. Pyrodictium and Methanopyrus species) have been found growing to 110°C [Stetter et al. (1983); Huber et al. (1989)]. So, in addition to their evolutionary significance, the archaebacteria represent a source of "robust" molecules which may possess unique capabilities of value to current and future biotechnology [Hough & Danson (1989)].

1.4. Thermoplasma acidophilum

1.4.1. Classification of Tp.acidophilum

Tp.acidophilum is a thermoacidophilic archaebacterium. First isolated from burning coal-refuse piles by Darland et al. (1970), it has since been

isolated from naturally occurring hot springs and sulfataric fields [Segerer et al. (1988)]. It lives at temperatures of between 50-64°C and at pH 0.5-3.0. (Table 1.4 summarises the main characteristics of the Thermoplasma genus.)

Tp.acidophilum cells are wall-less; as a result, the organism was originally classified along with the mycoplasmas [Darland et al. (1970)]. However, on the basis of 16S rRNA sequence comparisons, Tp.acidophilum has since been shown to belong to the methanogen-halophile branch of the archaebacteria (Fig. 1.5) [Yang et al. (1985)], whereas the classical mycoplasmas are believed to belong to the eubacteria [Fox et al. (1980)].

Although, on the basis of its rRNA, Tp.acidophilum appears to belong to the methanogen-halophile branch of the archaebacterial tree, phenotypically it resembles the sulphur-dependent thermophiles. In addition, its RNA polymerase subunit structure [Prangishvilli et al. (1982)] and functional analysis of its ribosome [Amils et al. (1989)] would seem to suggest that Tp.acidophilum is more closely related to the sulphur-dependent thermophiles than to the methanogens or halophiles. However, whereas other archaebacteria have linked rRNA genes, typically in the order 5'16S-23S-5S 3' [Neumann et al. (1983)], Tp.acidophilum is unique in possessing unlinked rRNA genes [Tu & Zillig (1982)]. These molecular contradictions suggest that the genus of Thermoplasma should perhaps be considered as a link between the two main branches of the archaebacterial tree, and may even prove to be a sole representative of a third branch.

1.4.2. Tp.acidophilum and its adaptation to an extreme environment

Tp.acidophilum, as a thermoacidophile, must possess numerous adaptations in order to live at moderately high temperatures and very low pH values. Some features of Tp.acidophilum, possibly relating to its thermal and

Table 1.4: Characteristics of members of the genus Thermoplasma

Species	Morphology	Metabolism	Temp opt(°C)	pH opt	G+C (mol%)	Genome size(kB)
Tp.acidophilum	irregular	obligately heterotrophic; uses S when growing anaerobically	59	1-2	46	1.4 x 10^6
Tp.volcanium	"	"	60	2	38	ND

Data from Segerer et al. (1988).

ND: not determined.

acidic adaptation, will now be discussed.

Tp.acidophilum grows at between pH 0.5-3.0 yet it maintains its internal cytoplasmic pH at approximately pH 6. Other acidophiles (including archaebacteria) regulate their internal pH by employing an H^+-translocating ATPase. However, this enzyme has been reported to be absent in Tp.acidophilum, and the organism appears instead to rely on an electron transport chain for maintenance of its internal pH [Searcy & Whatley (1982)]. The environmental conditions of Tp.acidophilum would therefore appear to be too extreme for an H^+-translocating ATPase [Searcy (1986)].

Tp.acidophilum membrane lipids are ether-linked C_{40} isopranyl glycerolipids [Langworthy et al. (1972)] typical of other archaebacterial lipids, and atypical of eukaryotic and eubacterial ester-linked lipids. Some thermophilic eubacteria, for instance Thermotoga maritima [Huber et al. (1986)], Thermodesulfotobacterium commune [Langworthy et al. (1983)], and Thermomicrobium roseum [Pond et al. (1986)] have similarly "unconventional" membrane lipids. The presence of these novel lipids in Tp.acidophilum (as in other archaebacteria and a few thermophilic eubacteria) may well reflect an adaptation to living under extreme conditions.

Tp.acidophilum DNA has an average G-C base composition of 46% [Christiansen et al. (1975); Searcy & Doyle (1975)] and so its DNA does not gain thermal stability through having a prevalence of guanines and cytosines. A histone-like protein (HTa) that binds DNA, similar to eukaryotic histones in terms of basicity and amino acid composition, has been identified and purified from Tp.acidophilum [Searcy (1975); DeLange et al. (1981a)]. The function of this HTa appears to be stabilisation of the Tp.acidophilum DNA against thermal denaturation [Stein & Searcy (1978)].

Sequences (obtained by amino acid sequencing) are available for two Tp.acidophilum proteins: the HTa protein (discussed above) [DeLange et al. (1981b)] and ferredoxin

[Wakabayashi et al. (1983)]. In addition, the amino acid sequence of Tp.acidophilum glucose dehydrogenase has recently been derived from the gene sequence [Bright et al. (1990)]. However, simple sequence comparisons (especially those involving sequences obtained from evolutionarily divergent organisms) generally fail to allow the elucidation of the structural features that confer thermostability, since the addition of just a single hydrogen bond or salt bridge can result in a significant increase in thermal stability of a protein. Instead, three-dimensional (3D) structural comparisons and/or site-directed mutagenesis studies are required. Such studies have yielded some information on aspects of protein thermostability, and will be discussed further in Chapter 5.

1.4.3. Tp.acidophilum citrate synthase

To extend our knowledge of citrate synthase and in particular of the archaebacterial enzyme, Danson et al. (1985) initiated studies on Tp.acidophilum citrate synthase. Table 1.5 shows the kinetic parameters of Tp.acidophilum citrate synthase in relation to other archaebacterial citrate synthases [Danson et al. (1985)].

Smith et al. (1987) purified citrate synthase from Tp.acidophilum and showed the enzyme to be a dimer with subunit M_r 43 000 (± 2 000). In addition, the first sixteen amino acids of the N-terminus (presented in Fig. 3.1) were determined [Smith et al. (1987)].

1.5. Aims of this project

At the start of this project, it was apparent that protein sequences from archaebacteria were required to corroborate rRNA phylogenetic studies. Citrate synthase seemed an excellent candidate from which to obtain archaebacterial sequence data since the tertiary structure of the pig enzyme and a large number of non-archaebacterial citrate synthase sequences were available. The enzyme from Tp.acidophilum was chosen for

Table 1.5: Kinetic properties of archaebacterial citrate synthases

Organism	Temperature of assay (°C)	Catalytic activity			Inhibition		
		Specific activity (nmol/min per mg)	K_m-acetyl-CoA (μM)	K_m-oxaloacetate (μM)	K_i-ATP (mM)	K_i-NADH (mM)	K_i-2-oxoglutarate (mM)
Archaebacteria							
H. halobium	30	51	211	58	5.6	>10	not inhibited
H. volcanii	30	61	75	38	1.9	>10	>20
Nb. gregoryi	30	14	108	63	4.2	3.2	>20
Nb. pharaonis	30	26	446	210	>30	5.2	not inhibited
S. acidocaldarius	55	280	10	20	0.9	4.6	2.2
T. acidophilum	55	172	6	5	2.2	5.4	5.4
M. barkeri	30	2	7	18	<0.3	<0.1	<0.3

Reproduced from Danson et al. (1985).

study as the N-terminal sequence of this organism's citrate synthase was known.

Obtaining sequence data from Tp.acidophilum, in addition to extending our knowledge of archaebacteria in general, was seen as valuable in determining the true phylogenetic position of this organism within the archaebacterial tree. Moreover, studying citrate synthase from the thermophilic Tp.acidophilum was attractive in terms of the potential elucidation of factors conferring thermostability to proteins.

The experimental aims of the project were:

(1) to clone and sequence the gene encoding citrate synthase of Tp.acidophilum;
(2) to extend citrate synthase sequence comparisons;
(3) to express the Tp.acidophilum citrate synthase gene in E.coli;
(4) to purify Tp.acidophilum citrate synthase expressed in E.coli.

CHAPTER 2: MATERIALS AND METHODS

2.1. Materials

2.1.1. Enzymes, reagents and other materials

Restriction endonucleases; DNA-modifying enzymes; 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (X-gal); isopropyl-ß-D-thiogalactoside (IPTG); and Lambda DNA-HindIII molecular weight markers were from Northumbria Biochemicals Limited, Cramlington, UK. Coenzyme A; 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB); oxaloacetate; ampicillin; and agarose (standard and low melting point) were from Sigma, Poole, UK. Bradford reagent was from Bio-Rad Laboratories, Munchen, West Germany. [γ-^{32}P]ATP (5000 Ci/mmol); [α-^{35}S]dATP (1000 Ci/mmol); and Hybond N hybridisation membrane were from Amersham International, Bucks., UK. NEN GeneScreen Plus hybridisation transfer membrane and NENsorb (nucleic acid purification) columns were from DuPont, Stevenage, UK. Sequenase kits were from United States Biochemical Corporation, Ohio, USA. X-ray film was from Fuji Photo Film Company, Japan. Acrylamide sequagel kits were from National Diagnostics, New Jersey, USA. Electran grade acrylamide and N,N,N',N'-tetramethylethylene diamine (TEMED) were from BDH, Poole, UK. Polyallomer Quick-seal centrifuge tubes were from Beckman Instruments Inc., Palo Alto, USA. Polysulfone filters were from Millipore, Massachusetts, USA. Sepharose CL-6B, polybuffer and protein molecular weight standards were from Pharmacia, Uppsala, Sweden. Redistilled phenol was from Rathburn Chemicals Ltd., Walkerburn, UK. Tryptone, yeast extract and agar were from Difco, Michigan, USA.

All other chemicals were from BDH, Fisons and Sigma (all UK). All other solvents (standard laboratory grade) were from BDH, UK and Fisons, Loughborough, UK.

Oligonucleotide probes and sequencing primers were synthesised on an Applied Biosystems 381A DNA synthesiser using phosphoramidite chemistry.

2.1.2. Bacterial strains, culture conditions and plasmids

A Thermoplasma acidophilum (DSM1728) freeze-dried culture (obtained from Deutsche Sammlung von Mikroorganismen, Braunschweig, West Germany) was used. The culture was grown in 1 litre of medium (described below) at 55°C and shaken at 200 rpm. The medium was as follows: $(NH_4)_2SO_4$ 1.3 $g.l^{-1}$, KH_2PO_4 0.28 $g.l^{-1}$, $MgSO_4.7H_2O$ 0.25 $g.l^{-1}$, $CaCl_2.2H_2O$ 70 $mg.l^{-1}$, $FeCl_3.6H_2O$ 20 $mg.l^{-1}$, $MnCl_2.4H_2O$ 1.8 $mg.l^{-1}$, $Na_2B_4O_7.10H_2O$ 4.5 $mg.l^{-1}$, $ZnSO_4.7H_2O$ 0.22 $mg.l^{-1}$, $CuCl_2.2H_2O$ 0.05 $mg.l^{-1}$, Na_2MoO_4 0.03 $mg.l^{-1}$, $VOSO_4.H_2O$ 0.03 $mg.l^{-1}$, $CoSO_4.6H_2O$ 0.01 $mg.l^{-1}$, yeast extract 1.0 $g.l^{-1}$. The pH was adjusted to pH 2.0 at RT with concentrated H_2SO_4. (This was pH 1.5 at 55°C.)
A 20 ml volume of 0.5 $g.ml^{-1}$ glucose (autoclaved separately) was added at inoculation time to give a final glucose concentration of 1% (w/v).

The E.coli strains DH5α: supE44 ΔlacU169 (ϕ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1; and TG1: supE hsdΔ5 thi Δ(lac-proAB) F'[traD36 proAB$^+$ lacI^q lacZΔM15] were used.

E.coli cultures were grown routinely in 2x YT medium (20 $g.l^{-1}$ tryptone, 10 $g.l^{-1}$ yeast extract and 10 $g.l^{-1}$ NaCl). Defined medium for E.coli consisted of an autoclaved solution of 10.5 $g.l^{-1}$ dipotassium hydrogen phosphate, 4.5 $g.l^{-1}$ potassium dihydrogen phosphate, 1 $g.l^{-1}$ ammonium sulphate, 0.5 $g.l^{-1}$ sodium citrate. Filter-sterilised solutions of the following were then added to give final concentrations of 0.2 $g.l^{-1}$ magnesium sulphate heptahydrate, 2.0 $g.l^{-1}$ carbon source and 5 $\mu g.l^{-1}$ of required amino acids.

Strains carrying plasmids conferring ampicillin resistance were propagated with 100 $\mu g.ml^{-1}$ ampicillin.

For solid medium and top agar, 1.5% (w/v) and 0.8% (w/v) agar was used, respectively.

Glycerol was added to cultures stored at -20°C and -70°C to a final concentration of 20% (v/v).

pUC (Ap^r), Bluescript (Ap^r) and M13 vectors were from Northumbria Biochemicals Limited, UK.

2.2. Methods

2.2.1. Electrophoresis conditions

DNA was analysed routinely on horizontal agarose gels. Agarose solutions were made in 1x TBE (90 mM Tris-base, 90 mM boric acid and 2 mM EDTA) to the desired concentration. Ethidium bromide was added to a concentration of 0.5 $\mu g.ml^{-1}$. Gels were run buffered in 1x TBE containing 0.5 $\mu g.ml^{-1}$ ethidium bromide at approximately 5 V/cm. A 0.2x volume of loading buffer (15% (w/v) Ficoll, 0.25% (w/v) bromophenol blue and 0.25% (w/v) xylene cyanol FF) was added to the sample prior to loading. HindIII-digested Lambda DNA fragments were used as molecular weight markers.

Polyacrylamide urea sequencing gels (6-8%, w/v;) were made according to instructions supplied with the Sequagel kit and assembled by a sliding plate technique described by Williams et al. (1986). Gels (length 40cm and thickness 2 mm) were run buffered in 1x TBE at 1500 V for 2-6 h.

Polyacrylamide gel electrophoresis (PAGE) in 0.1% (w/v) SDS was carried out with 10% (w/v) polyacrylamide gels. The resolving gel consisted of 10 ml 30% (w/v) acrylamide: 0.8% (w/v) bisacrylamide solution, 3.75 ml 3 M Tris-HCl (pH 8.8), 0.3 ml 10% (w/v) SDS, 15.6 ml H_2O, 0.15 ml 15% (w/v) ammonium persulphate and 15 μl TEMED. The stacking gel consisted of 1.25 ml 30% (w/v) acrylamide: 0.8% (w/v) bisacrylamide solution, 5 ml 0.5 M Tris-HCl (pH 6.8), 0.1 ml 10% (w/v) SDS, 6.1 ml H_2O, 50μl 15% (w/v) ammonium persulphate and 5 μl TEMED. An equal volume of loading buffer [0.125 M Tris-HCl (pH 6.8), 4% (w/v) SDS, 10% (v/v) ß-mercaptoethanol, 20% (w/v) sucrose, 0.004% (w/v) bromophenol blue] was added to samples; these were then boiled for 2-3 min, immediately prior to loading. Gels (length 14 cm and thickness 1 mm) were run in 25 mM Tris-HCl (pH 8.3), 0.19 M glycine and 1% (w/v) SDS at 50V for 12 h. They were stained in a solution

containing 2.5 $g.l^{-1}$ coomassie Blue R, 9.2% (v/v) glacial acetic acid, 45.4% (v/v) methanol, and destained in a 10% (v/v) glacial acetic acid and 30% (v/v) methanol solution. Protein standards, consisting of phosphorylase b (M_r = 94 000), bovine serum albumin (M_r = 67 000), ovalbumin (M_r = 45 000), carbonic anhydrase (M_r = 30 000), trypsin inhibitor (M_r = 20 100) and lactalbumin (M_r = 14 400), were used.

2.2.2. Phenol/chloroform extraction of nucleic acid

Phenol and/or chloroform extractions were used routinely in order to purify DNA from a variety of sources. An equal volume of 50 mM Tris-HCl (pH 8.0)-saturated phenol:chloroform:isoamyl alcohol (24:24:1, v/v/v) was added to the sample. The mixture was vortexed for 1 min and then separated into two phases by centrifugation (10 000g at 10°C for 5 min). The upper aqueous phase containing the nucleic acid was transferred to a clean tube and the phenol phase discarded.

2.2.3. Ethanol precipitation of nucleic acid

Ethanol precipitations were used routinely to concentrate DNA and/or to change the solvent in which it was dissolved.

A one-tenth volume of 3 M sodium acetate (pH 5.5) (prepared by bringing a 3 M solution to pH 5.5 with glacial acetic acid) and 2-3 volumes of ice-cold absolute ethanol were added to the sample. The DNA was precipitated at -20°C for 30 min and then pelleted by centrifugation (10 000g at 4°C for 30 min). The supernatant was discarded, the nucleic acid pellet rinsed with 70% (v/v) ice-cold ethanol and resedimented as before. The nucleic acid was vacuum-dried in a lyophiliser (Speed Vac Concentrator, Savant Instruments Inc., UK) and resuspended in the desired volume of TE (10 mM Tris-HCl, 1 mM EDTA) buffer (pH 8.0).

2.2.4. Purification of nucleic acid on a NENsorb column

The NENsorb column was washed with 2 ml 100% methanol and then with 2 ml Buffer A [0.1 M Tris-HCl (pH 7.7), 10 mM triethylamine (TEA), 1 mM EDTA]. The nucleic acid was loaded onto the column in a volume of 0.3-1.0 ml 10 mM TEA. The column containing the bound nucleic acid was rinsed with 3 ml Buffer A and then with 0.5 ml H_2O. The nucleic acid was eluted with 0.5 ml 50% (v/v) methanol, dried in a vacuum and resuspended in the desired volume of TE buffer (pH 8.0).

2.2.5. Quantitation of DNA

The concentration of the DNA was determined by carrying out an OD_{260}/OD_{280} analysis. An absorbance of 1 at wavelength 260 nm corresponds to approximately 37 $\mu g.ml^{-1}$ single-stranded DNA or 50 $\mu g.ml^{-1}$ double-stranded DNA. An OD_{260}/OD_{280} ratio of greater than 1.8 was taken to indicate that the sample was pure.

2.2.6. Digestion of DNA by restriction endonucleases

DNA was digested with the desired enzyme under the buffer conditions defined for that particular enzyme. One unit of enzyme activity is defined as the amount of enzyme that will cut at all its specific sites in a 1 μg sample of lambda DNA in 1 h at 37°C. In order to ensure complete digestion of DNA, excess units of enzyme were used. Enzyme reactions were stopped by the addition of EDTA (pH 8.0) to a final concentration of 10 mM and/or by incubation at 65°C. When the DNA was digested with two enzymes which required different buffer conditions, it was incubated first for 1 h with the enzyme that required the lower salt concentration; the second enzyme was then added along with an appropriate volume of NaCl.

2.2.7. Treatment of DNA with S1 nuclease

DNA was incubated with 10 units S1 nuclease in 0.5 ml S1 buffer [33 mM sodium acetate (pH 5.0), 50 mM NaCl, 0.03 mM $ZnSO_4$] at 37°C for 30 min. The

reaction was terminated by a phenol/chloroform extraction.

2.2.8. Preparation of Tp.acidophilum genomic DNA: Method I

A litre culture of Tp.acidophilum was grown at 55°C as described in Materials (Section 2.1.2). Approximately 0.2 g cells were harvested by centrifugation (5 000g at 4°C for 15 min) and resuspended in 60 ml 0.2 M Tris-HCl (pH 8.0) containing 20 mM EDTA. The suspension was incubated on ice for 30 min in order to allow complete lysis of the cells to occur. Cell debris was removed by centrifugation (7 000g at 4°C for 15 min). Extractions with phenol, then with phenol-chloroform (1:1, v/v) and finally with chloroform were carried out. The total nucleic acid was precipitated with 2 volumes of ice-cold absolute ethanol, pelleted by centrifugation (10 000g at 4°C for 20 min), and resuspended in 10 ml TE buffer (pH 8.0). CsCl to a concentration of 1 $g.ml^{-1}$, and ethidium bromide to a concentration of 0.1 $mg.ml^{-1}$ were added. The solution was transferred to a polyallomer quick-seal tube and centrifuged in a Ti50 rotor (80 000g for 48 h). The DNA was recovered and extracted several times with equal volumes of CsCl-saturated-propan-2-ol in order to remove the ethidium bromide. The solution was diluted with a 2x volume of distilled H_2O. The DNA was then precipitated with ethanol (at RT), pelleted and finally resuspended in 1 ml TE buffer (pH 8.0).

2.2.9. Preparation of Tp. acidophilum genomic DNA: Method II

Approximately 0.2 g Tp.acidophilum cells were harvested from a litre of freshly grown culture by centrifugation (5 000g at 4°C for 15 min) and resuspended in 3.2 ml 50 mM Tris-HCl buffer (pH 8.0) containing 25% (w/v) sucrose. A 2 ml volume of a 5% (w/v) SDS solution containing 0.15 M EDTA (pH 8.0) was added to the cell suspension, and the mixture swirled on ice for

5 min. Proteinase K (2.5 mg) was added and the cell lysate incubated at 45°C for 30 min. A one-tenth volume of 3 M sodium acetate (pH 5.5) was added and a phenol:chloroform (1:1, v/v) extraction carried out. Total nucleic acid was precipitated with 2 volumes of ice-cold absolute ethanol. The precipitate was spooled using a glass rod and resuspended in 5 ml of 2x SSC (0.3 M NaCl; 0.03 M sodium citrate). A 0.25 ml volume of 10 $mg.ml^{-1}$ RNAse was added and the mixture incubated at 37°C for 20 min. The mixture was extracted with an equal volume of phenol:chloroform (1:1, v/v). The DNA was precipitated with 2.5 volumes of ice-cold ethanol and pelleted by centrifugation (10 000g at 4°C for 30 min). The DNA pellet was dried briefly in a vacuum and then resuspended in 4.5 ml distilled water. This last ethanol precipitation step was repeated. The DNA was finally resuspended in 2 ml distilled water.

2.2.10. Ammonia-treatment of oligonucleotides

The column with newly synthesised oligonucleotide attached was placed between two 1 ml syringes, one of which contained 1 ml of concentrated ammonium hydroxide solution. Initially, 0.3 ml of the ammonium hydroxide were sucked into the column; then, after 20 min, further 0.2 ml aliquots were drawn through at 20 min intervals until all the solution had been passed through. The crude DNA/ammonium hydroxide solution was heated at 55°C overnight. The sample was freeze-dried to remove the ammonia and resuspended in 1 ml distilled H_2O. The oligonucleotide was then precipitated with ethanol (as described in section 2.2.3) and finally resuspended in 0.5 ml distilled H_2O.

2.2.11. End-labelling

Approximately 20 pmol of oligonucleotide were incubated with 100 μC_i [γ-^{32}P] ATP, 1x kinase buffer [50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$] and 10 units of T_4 kinase in a total volume of 10 μl at 37°C for 20 min.

The labelled probe was then separated from unincorporated [γ-^{32}P] ATP by passage through a NENsorb column (see Section 2.2.4).

2.2.12. Autoradiography

Autoradiography with [γ-^{32}P] labelled DNA was carried out at -70°C using pre-flashed X-ray film and intensifying screens. Autoradiography with [α-^{35}S] was carried out at room temperature using unflashed X-ray film.

2.2.13. Southern blotting of Tp.acidophilum DNA

In a Southern blot, DNA is transferred from a gel on to a membrane and probed. In this study a capillary method, based on that of Southern (1975), was employed and is outlined below.

Aliquots of Tp.acidophilum DNA (2.5 μg) were restricted to completion with one and/or two hexanucleotide-specific restriction endonucleases. The digests were run on a 0.8% (w/v) agarose gel at 3 V/cm for 12-16 h.

The gel was incubated in 0.25 M HCl for 15 min at RT and then transferred to a 0.4 M NaOH solution containing 0.6 M NaCl for 30 min at RT. It was washed twice in a solution containing 1.5 M NaCl and 0.5 M Tris-HCl (pH 7.5) for 15 min at RT.

A piece of GeneScreen Plus was cut to the exact size of the gel and the concave side (designated side B) was marked. The membrane was soaked in distilled water for 15 min and then transferred to a 10x SSC solution for 10 min.

The gel, inverted-side-up, was placed on a stand and the pre-wetted membrane, side B facing down, was placed on top. Four sheets of 3MM and a two inch layer of paper, all cut to the same size as the gel and the membrane, were layered above. A plastic cover and a 1 kg weight were then placed on top of the paper. The wells, on either side of the stand, were filled with 10x SSC. The blot was then left to run for 16-24 h.

The membrane was lifted off the gel and immersed briefly (30-60 s) in a 0.4 M NaOH solution. It was transferred to a solution containing 0.2 M Tris-HCl (pH 7.5) and 2x SSC for 10 min at RT. Finally it was dried by incubation at RT for 15 min and at 37°C for 15 min.

The membrane was prehybridised in approximately 10 ml hybridisation solution [50 mM Tris-HCl (pH 7.6), 0.9 M NaCl, 1% (w/v) SDS] in a sealed polythene bag at 37°C for 6 h. [γ-^{32}P] end-labelled oligonucleotide and 5 mg sonicated salmon sperm DNA were boiled for 2-3 min and added to the bag containing the membrane. Hybridisation was carried out at the desired hybridisation temperature (T_h) for 16 h. The T_h was estimated according to the following formula: T_h = 2°C x (number of A + T residues) + 4°C x (number of G + C residues) - 5°C [Wallace et al. (1979)].

After an initial rinse in 6x SSC at RT for 5 min, the hybridised membrane was washed in 6x SSC containing 1% (w/v) SDS at the desired temperature for 15 min. The washes were made more stringent by increases of temperature. The membrane was monitored with a Geiger counter and autoradiographed at suitable intervals. Washing of increased stringency was continued until background had been reduced to a desirable level.

2.2.14. Extraction of DNA from agarose

There are many methods that can be used to extract DNA from agarose. In this study, a freeze-squeeze method based on that of Tautz & Renz (1983) was found to give adequate yields of good quality DNA for cloning. This method is outlined below.

The restricted DNA was run on a low melting point (LMP) agarose gel. The appropriate band was excised and placed in a 0.5 ml eppendorf which had been pierced and plugged with siliconised glass wool. After freezing in liquid nitrogen for 1 min, the eppendorf containing the section of agarose was placed inside a 1.5 ml eppendorf

and spun (10 000g at RT for 15 min). The solution in the outer 1.5 ml eppendorf contained the DNA, whereas the inner 0.5 ml eppendorf still contained the agarose. The latter tube was discarded.

DNA, recovered by this freeze-squeeze method, was then purified by passage through a NENsorb column.

2.2.15. Ligation of DNA

Ligation was carried out with a vector to insert ratio of five to one in 20 μl ligation buffer [50 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, 100 $\mu g.ml^{-1}$ BSA] with 1 unit of T_4 ligase at 15°C for 4-16 h. Amounts in the order of 50-100 ng vector were used.

2.2.16. Preparation of competent E.coli cells

A volume of 0.5 ml of an overnight culture of E.coli cells was used to inoculate 50 ml 2x YT broth. This culture was incubated at 37°C with agitation till an OD_{550} of 0.3-0.5 was attained. It was then chilled on ice for 10-15 min. Cells were pelleted by centrifugation (3 000g at 4°C for 5 min) and resuspended in 16.5 ml sterile 50 mM $CaCl_2$ solution. After incubation on ice for 30 min, the cells were pelleted by centrifugation (2 000g at 4°C for 5 min) and resuspended in 4 ml sterile 50 mM $CaCl_2$.

Optimum transformation frequencies were obtained from freshly-made competent cells; however, it was found that cells prepared as above remained competent for up to 24 h.

2.2.17. Transformation of E.coli with plasmid DNA

The desired quantity of DNA (when transforming ligated material, approximately 0.1x volume of the ligation mix) was added to a 0.2 ml aliquot of competent E.coli cells. The transformation mixture was incubated on ice for 30 min and then heat-shocked at 42°C for 90 s. A 0.8 ml volume of 2x YT was added and the mixture

incubated at 37°C for 1 h. The transformed material (0.1 - 0.2 ml per plate) was then plated out onto selective 2x YT agar.

If a plasmid with the ß-galactosidase colour detection system was being used, 50 µl 2% (w/v) X-gal and 10 µl 100 mM IPTG were added to the transformation mix immediately prior to its being plating out.

2.2.18. Transformation with replicative form (RF) M13 DNA

The desired quantity of DNA (when transforming ligated material, the whole ligation mix) was added to a 0.2 ml aliquot of competent E.coli cells. The transformation mixture was incubated on ice for 30 min. It was then mixed with 0.2 ml of exponentially-growing E.coli cells, 0.3 ml 2x YT top agar, 50 µl 2% (w/v) X-gal and 20 µl 100 mM IPTG, and poured on to 2x YT agar plates.

2.2.19. Colony lifts

Colony lifts, using a method based on that of Grunstein and Hogness (1975), were carried out as described below.

Hybond N membrane was cut to the size of the plates to be blotted, marked with a non-symmetrical pattern and lowered on to the agar surface. After 1 min, the membrane was removed and placed colony-side-up on a pad of filter paper, which had been soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH). After 7 min, the membrane was transferred to a pad of filter paper, which had been soaked in neutralising solution [1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2), 1 mM EDTA]. After 3 min, the membrane was transferred to fresh pads, which had been soaked in neutralising solution for a further 3 min. The membrane was washed briefly in 2x SSC and then dried at RT. The DNA was fixed to the membrane by baking at 80°C for 2 h. The colony blot was then prehybridised, hybridised and washed as described for Southern blots in Section 2.2.13.

2.2.20. Rapid preparation of plasmid DNA ("miniprep")

The following protocol describes how plasmid can be prepared in small quantities for analysis. (Plasmid produced by this method was not suitable for sequencing.)

Cells from an overnight culture (1.5 ml) of the strain containing the desired plasmid were harvested by centrifugation (5 000g at 4°C for 10 min) and resuspended in a solution containing 50 mM glucose, 25 mM Tris-HCl (pH 8.0) and 10 mM EDTA. Lysozyme was added to a concentration of 2 $mg.ml^{-1}$ and the mixture incubated at RT for 10 min. A 0.2 ml volume of a 0.2 M NaOH solution containing 1% (w/v) SDS was added and the mixture incubated on ice for 10 min. A 0.15 ml volume of 3 M sodium acetate (pH 5.5) was added and the mixture incubated on ice for 30 min. Cell debris was pelleted by centrifugation (10 000g at 4°C for 10 min) and the supernatant transferred to a clean eppendorf. Two phenol/chloroform extractions followed by an ethanol precipitation were carried out. The nucleic acid pellet was resuspended in 20 μl TE buffer (pH 8.0). Finally, 1 μl 10 $mg.ml^{-1}$ RNAse was added and the mixture incubated at 65°C for 10 min.

2.2.21. Large-scale plasmid preparation ("maxiprep")

Cells from 500 ml of an overnight culture of the strain carrying the desired plasmid were harvested by centrifugation (5 000g at 4°C for 10 min). They were resuspended in 8 ml of a solution containing 50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA and 5 $mg.ml^{-1}$ lysozyme. The mixture was transferred to corex tubes and incubated at RT for 15 min. A 16 ml volume of a 0.2 M NaOH solution containing 1% (w/v) SDS was added and the mixture incubated on ice for 10 min. A 12 ml volume of 5 M potassium acetate (pH 5.5) was added and the mixture incubated on ice for 15 min. Cell debris was pelleted by centrifugation (10 000g at 4°C for 40 min). The supernatant was transferred to a clean tube and precipitated with 0.6 volumes of isopropanol at RT for

15 min. The nucleic acid was pelleted by centrifugation (10 000g at 10°C for 10 min) and resuspended in 9.5 ml distilled H_2O. A 0.3 ml volume of 10 $mg.ml^{-1}$ ethidium bromide and 9.5 mg CsCl were added to the solution, which was then transferred to a polyallomer quick-seal tube and centrifuged in a Ti50 rotor (80 000g for 48 h). Generally two plasmid bands were obtained on such a CsCl gradient; the top band corresponding to nicked circular plasmid DNA and the lower band corresponding to supercoiled plasmid DNA. The lower band was removed and the DNA purified as described in Section 2.2.8. It was resuspended in 0.5 ml TE buffer (pH 8.0).

2.2.22. Preparation of single-stranded M13 sequencing template

A fresh, exponential culture of E.coli TG1 cells was diluted 1:100 in 1.5 ml of 2x YT medium and infected with an M13 plaque. The culture was shaken vigorously at 37°C for 6-8 h. Cell debris was pelleted by centrifugation (10 000g at 4°C for 10 min) and the phage-containing supernatant transferred to a fresh tube. A one-ninth volume of 40% (w/v) PEG-8000 solution containing 5 M sodium acetate (pH 7.0) was added and the mixture incubated at 4°C for 15 min. The phage-precipitate was pelleted by centrifugation (10 000g at 4°C for 15 min) and resuspended in 0.2 ml TE buffer (pH 8.0). Two phenol extractions and an ethanol precipitation were carried out. The DNA was finally resuspended in 20 µl TE buffer (pH 8.0). (A volume of 7 µl - containing approximately 1 µg DNA - was generally enough for one sequencing reaction.)

2.2.23. Preparation of single-stranded sequencing template from double-stranded vector

A Sepharose CL-6B column (in a 0.5 ml eppendorf pierced, and plugged with glass beads) was packed by centrifugation in a swing-out rotor (3 000g at RT for

4 min). The column was washed twice with 50 µl TE buffer (pH 8.0) by centrifugation (3 000g at RT for 4 min). A 20 µl (10 µg) volume of "maxiprep" plasmid DNA was denatured by the addition of 5 µl 1 M NaOH solution containing 1 mM EDTA, and incubation at RT for 5 min. The DNA was neutralised by passage through the prepared column (3 000g at RT for 4 min). A 7 µl volume of this denatured DNA was used for one sequencing reaction.

2.2.24. Sequencing

Sequencing was carried out using the dideoxynucleoside chain termination method of Sanger et al. (1977) as outlined below.

The primer (0.5 pmol) was annealed to single-stranded template (7 µl), in a solution containing 40 mM Tris-HCl (pH 7.5), 20 mM $MgCl_2$ and 50 mM NaCl, by incubation at 65°C for 2 min followed by a gradual cooling to RT.

Volumes of 2.5 µl of the four termination mixes - ddGTP, ddATP, ddCTP and ddTTP (ddX termination mix consisting of 80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddXTP and 50 mM NaCl) - were aliquotted into four separate eppendorfs and incubated at 37°C for at least 1 min.

A solution containing 1 µl 0.1 M DTT, 2 µl labelling mix (1.5 µM dGTP, 1.5 µM dCTP 1.5 µM dTTP), 5 μC_i [α-^{35}S]dATP and 2 µl 8-fold diluted Sequenase was added to the annealed primer-template solution, and the mixture was incubated for 2 min at RT. Aliquots of 3.5 µl of this mixture were then added to each of the four termination mixes. The tubes were incubated at 37°C for 5 min. Finally, 4 µl of stop solution [95% (v/v) formamide, 20 mM EDTA, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol FF)] were added to each tube.

The samples were heated to 75°C for 2 min prior to loading on to polyacrylamide sequencing gels. A volume of 3 µl was used in each lane.

2.2.25. Assay for citrate synthase activity

Samples were assayed spectrophotometrically for Tp.acidophilum citrate synthase at 55°C according to the method of Srere et al. (1963). In this assay the increase in absorbance due to the production of the yellow-coloured thio-nitrobenzoate anion from DTNB is monitored at 412 nm (E_{412} 13 600 $l.mol^{-1}. cm^{-1}$). The assay mixture used in this study contained 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.2 mM acetyl CoA, 0.2 mM oxaloacetate and 0.1 mM DTNB. The acetyl CoA was prepared by cooling a 10 $mg.ml^{-1}$ CoA solution on ice, and then adding 0.2 ml 1 M $KHCO_3$ and 0.1 ml 1 M acetic anhydride.

2.2.26. Bradford protein determination

A 50 µl volume of sample was mixed with 25 µl 0.1 M NaOH, 0.725 ml TE (pH 8.0) buffer, and 0.2 ml BioRad Bradford reagent and incubated at RT for 10 min. The samples were assayed spectrophotometrically at 595 nm. Protein concentrations were determined from a calibration carried out with 5-25 µg BSA.

2.2.27. Production of Tp.acidophilum citrate synthase in E.coli

Typically a 10 ml culture of E.coli, containing recombinant plasmid which carried the Tp.acidophilum citrate synthase gene, was grown for approximately 16 h in 2x YT medium with ampicillin. For large scale production 500 ml cultures were grown.

If IPTG was used, it was added to a concentration of 10 mM approximately 2 h prior to completion of growth.

2.2.28. Preparation of cell-free extract from E.coli

Cells from 10 ml cultures were harvested by centrifugation (5 000g at 4°C for 10 min) and resuspended in 0.5 ml of 20 mM Tris-HCl (pH 8.0) containing 1 mM EDTA. The samples were sonicated with a 4.5 mm probe at 50 W for 1 min. Cell debris was removed by

centrifugation (10 000g at 4°C for 10 min).

Cells from 500 ml cultures were harvested by centrifugation (5 000g at 4°C for 10 min) and resuspended in 10 ml 20 mM Tris-HCl (pH 8.0) containing 1 mM EDTA, and sonicated with a 20 mm probe at 180 W for 1 min. Cell debris was removed by centrifugation (10 000g at 4°C for 10 min).

2.2.29. Heat purification

Cell-free extracts of E.coli were heated at the desired temperature for the desired length of time. Denatured protein was then removed by centrifugation (10 000g at 4°C for 10 min).

2.2.30. Chromatofocussing

Chromatofocussing was carried out on a Pharmacia FPLC Mono P (HR 5/20) column, previously equilibrated in 25 mM diethanolamine-HCl (pH 9.5). Protein was eluted by applying 10% [v/v] Polybuffer 96 (pH 6.0). (This generated a pH gradient of pH 9.0 to pH 6.0.) Fractions of 1 ml were collected.

Polybuffer retained after chromatofocussing was removed from the sample, by passing it through a polysulfone filter with an exclusion size of M_r 10 000 (2 000g at 4°C for 1 h). This step also served to concentrate the sample.

2.2.31. Computer-aided sequence alignment

The multiple citrate synthase computer-aided sequence alignment was carried out by Dr. C.M. Henneke according to Henneke et al. (1989). The program employed uses BESTMULT in conjunction with the McLachlan simscore matrix and gives the following penalties: eight for breaks within a helix or a sheet; two for breaks occurring in the one or two amino acids at the end of a helix or a sheet; and one for breaks in non-secondary structure. It has a gap length factor of 0.

2.2.32. Secondary-structure prediction

The secondary-structure prediction was carried out with the help of Dr. G. Taylor using a program package called PREDICT. This gives a joint prediction based on results obtained by eight different methods (Burgess; Dufton & Hider; Chou & Fasman; Garnier; Lim; McLachlan; Nagama; and Kabat) [Taylor (1990) and references therein].

CHAPTER 3: CLONING AND SEQUENCING OF *Thermoplasma acidophilum* DNA

3.1. Introduction

This chapter describes the work involved in the cloning of the *Tp.acidophilum* citrate synthase gene into *E.coli*.

3.1.1. Design of probes for the Tp.acidophilum citrate synthase gene

The N-terminal amino acid sequence available for the *Tp.acidophilum* citrate synthase enzyme [Smith *et al.* (1987)] was used as a basis for designing oligonucleotide probes for the *Tp.acidophilum* citrate synthase gene (Fig. 3.1). There are two criteria to consider when designing gene probes from amino acid sequence information, namely maximum length and minimum complexity (where complexity is due to wobble introduced as a result of the degeneracy of the genetic code). Complexity of the probe can be reduced by employing a "best guess" approach if the codon preference of the organism is known. However, at the start of this study, no protein-encoding genes had been cloned from *Tp.acidophilum*, and therefore there was no information relating to its codon usage.

Two probes were designed for the citrate synthase gene: a redundant 17-mer oligonucleotide (designated CS-17-er) based on the first six amino acids of the N-terminal sequence and taking into account all possible bases at each of the wobble positions (Fig. 3.1.b); and a non-redundant 48-mer oligonucleotide (designated CS-48-mer) based on all of the sixteen known amino acids of the N-terminal sequence (Fig. 3.1.c). Calculated guesses were made at each of the 20 wobble positions in this CS-48-mer, using codon usage data from the *Sulpholobus solfataricus* ß-galactosidase gene shown in Table 3.1 [Rossi (1988); Cubellis *et al.* (1990)].

Figure 3.1: The N-terminal amino acid sequence of Tp.acidophilum citrate synthase and sequences of oligonucleotides designed to probe for this sequence

(a)

N-terminal	P	E	T	E	E	I	S	K	G	L	E	D	V	N	I	K
Putative 5'ends of gene	CCA	GAA	ACA	GAA	GAA	ATA	TCA	AAA	GGA	CTC	GAA	GAC	GTA	AAC	ATA	AAA
	C	G	C	G	G	C	G	G	C	T	G	T	C	T	C	G
	G		G			T	AGC		G	TTA			G		T	
	T		T				T		T	G			T			

(b)

CS-17-mer	CCA	GAA	ACA	GAA	GAA	AT
	C	G	C	G	G	
	G		G			
	T		T			

(c)

CS-48-mer	CCA	GAA	ACT	GAG	GAA	ATC	AGT	AAA	GGA	CTC	GAG	GAT	GTT	AAT	ATC	AAA

Table 3.1: Codon usage of S.solfataricus ß-galactosidase gene

F TTT	13.	S TCT	6.	Y TAT	16.	C TGT	1.
F TTC	10.	S TCC	3.	Y TAC	18.	C TGC	0.
L TTA	18.	S TCA	9.	* TAA	1.	* TGA	0.
L TTG	2.	S TCG	0.	* TAG	0.	W TGG	17.
L CTT	4.	P CCT	4.	H CAT	7.	R CGT	1.
L CTC	1.	P CCC	3.	H CAC	7.	R CGC	0.
L CTA	11.	P CCA	18.	Q CAA	7.	R CGA	0.
L CTG	1.	P CCG	1.	Q CAG	2.	R CGG	0.
I ATT	6.	T ACT	9.	N AAT	19.	S AGT	7.
I ATC	3.	T ACC	2.	N AAC	11.	S AGC	6.
I ATA	13.	T ACA	6.	K AAA	11.	R AGA	19.
M ATG	10.	T ACG	2.	K AAG	12.	R AGG	12.
V GTT	17.	A GCT	11.	D GAT	25.	G GGT	13.
V GTC	1.	A GCC	3.	D GAC	7.	G GGC	3.
V GTA	9.	A GCA	8.	E GAA	19.	G GGA	20.
V GTG	4.	A GCG	5.	E GAG	11.	G GGG	4.

Derived from data of Rossi (1988).

3.1.2. Choice of cloning strategy

In a ligation reaction between vector and insert DNA, the vector tends to religate to itself at a greater frequency than it ligates to the insert DNA. This problem can be overcome if the vector is dephosphorylated. However, to make a good dephosphorylated vector a reliable method for removal of phosphatase enzyme is required. An alternative approach to solving the vector religation problem, directional cloning, was employed in this study. This approach uses vector that is cut with two different restriction endonucleases; consequently, it can no longer religate to itself.

3.2. Results

3.2.1. Preparation of Tp.acidophilum genomic DNA

DNA was initially prepared by a method involving purification on a CsCl gradient (Section 2.2.8). However, this DNA was found to be badly sheared and unsuitable for cloning; therefore, an alternative method (Section 2.2.9) based on that used by Bowen *et al.* (1988) for *Thermus thermophilus* was employed. A litre of *Tp.acidophilum* culture yielded approximately 0.4 mg of genomic DNA. When analysed on an agarose gel (Fig. 3.2), the DNA ran as a tight band at a size equivalent to uncut Lambda DNA, indicating that it was of high molecular weight. When digested with restriction endonucleases, the DNA cut easily to give distinctive and repetitive banding patterns, indicating that it was very pure.

3.2.2. Preparation of Southern blots of restricted Tp.acidophilum DNA

A series of single and/or double digests of *Tp.acidophilum* genomic DNA was carried out with the hexanucleotide-specific restriction endonucleases: *BamHI*, *EcoRI*, *HindIII* and *PstI*. The single digests were run on one gel (Fig. 3.3) and the double digests run on a second

Figure 3.2: Analysis of Tp.acidophilum genomic DNA on a 1.0% (w/v) agarose gel

Lane 1: Lambda DNA cut with HindIII;
Lane 2: Tp.acidophilum genomic DNA.

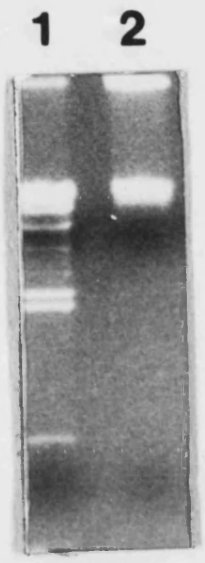
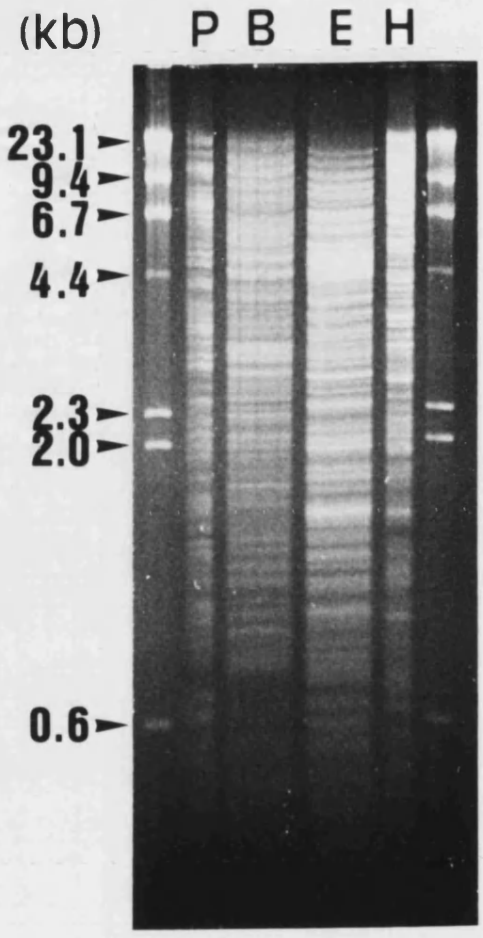

1
2

Figure 3.3: Tp.acidophilum genomic DNA digests separated on a 0.8% (w/v) agarose gel

B: BamHI; E: EcoRI; H: HindIII; P: PstI.

The sizes of the Lambda DNA-HindIII fragments (extreme left and right lanes) are indicated in the left-hand margin.


(kb)
P B E H
23.1
9.4
6.7
4.4
2.3
2.0
0.6

gel (Fig. 3.4). The two gels were blotted (separately) on to GeneScreen Plus, and the blots hybridised with the CS-17-mer and CS-48-mer oligonucleotides as described below.

3.2.3. Use of the CS-17-mer oligonucleotide

The two Southern blots described in the above section were hybridised with [γ-^{32}P] end-labelled CS-17-mer at 37°C. The blots were washed in 6x SSC containing 1% (w/v) SDS: twice at RT, twice at 37°C and twice at 42°C. The autoradiograph obtained for the blot of single enzyme digests (after its 42°C wash) is represented in Fig. 3.5 and the autoradiograph obtained for the blot of double enzyme digests (after its 37°C wash) is represented in Fig. 3.6.

DNA fragments to which the CS-17-mer hybridised are listed in the first column of Table 3.2. Even though single bands were not obtained for each digest, a putative restriction map of the region to which this probe hybridised is proposed in Fig. 3.7.

3.2.4. Cloning and sequencing of the 0.5-kb EcoRI-HindIII fragment identified by the CS-17-mer oligonucleotide

The 0.5-kb EcoRI-HindIII fragment, identified by the CS-17-mer, was selected for cloning because it is (generally) easier to clone, and obviously quicker to sequence, small fragments of DNA.

Approximately 25 μg Tp.acidophilum genomic DNA was digested with EcoRI and HindIII and separated on a 1.0% (w/v) low melting point (LMP) agarose gel. DNA of fragment-size approximately 0.5-kb was recovered by the freeze-squeeze method and purified on a NENsorb column. This size-selected DNA was then ligated into EcoRI-HindIII-cleaved Bluescript vector and transformed into competent E.coli TG1 cells. Approximately 300 transformants were transferred on to Hybond-N and hybridised with [γ-^{32}P] end-labelled CS-17-mer. The filter was washed in 6x SSC containing 1% (w/v) SDS: once

Figure 3.4: Tp.acidophilum genomic DNA digested with BamHI (B), EcoRI (E), HindIII (H) and PstI (P) in all possible double-digest combinations and separated on a 0.8% (w/v) agarose gel.

The sizes of the Lambda DNA-HindIII fragments (middle lane) are indicated in the left-hand margin.

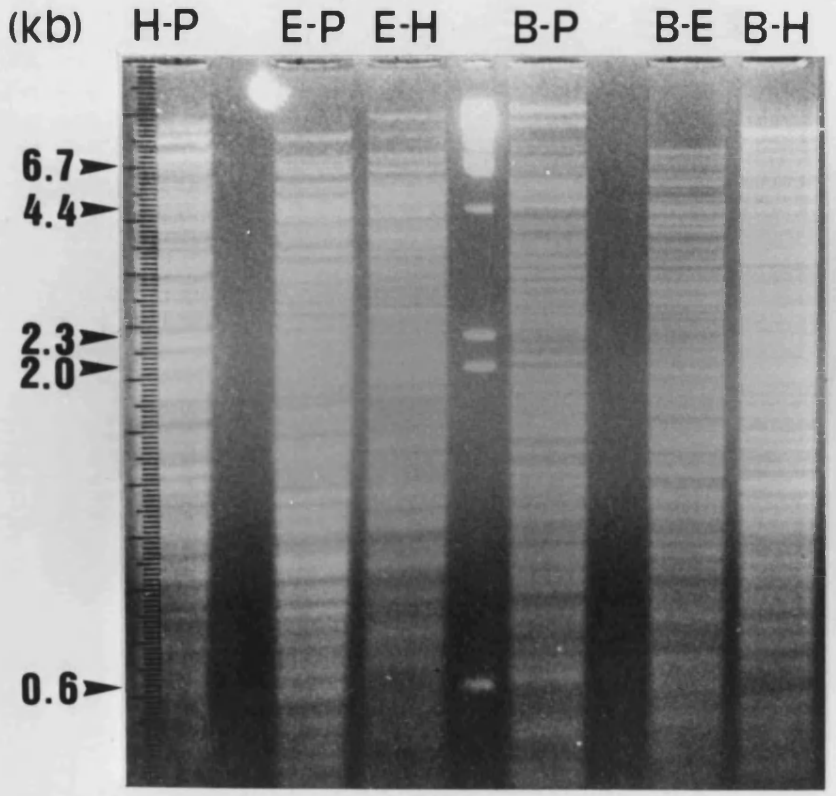

(kb)
H-P
E-P
E-H
B-P
B-E
B-H
6.7
4.4
2.3
2.0
0.6

Figure 3.5: Hybridisation of single enzyme digests of Tp.acidophilum DNA with end-labelled CS-17-mer oligonucleotide

The gel used for the blot shown here is described in Fig. 3.3. Experimental details are given in Section 3.2.3. Arrows in the left-hand margin indicate the migration points of Lambda DNA-HindIII fragments. The tracks are labelled according to the enzyme used for digestion of the Tp.acidophilum DNA.

B: BamHI; E: EcoRI; H: HindIII; P: PstI.

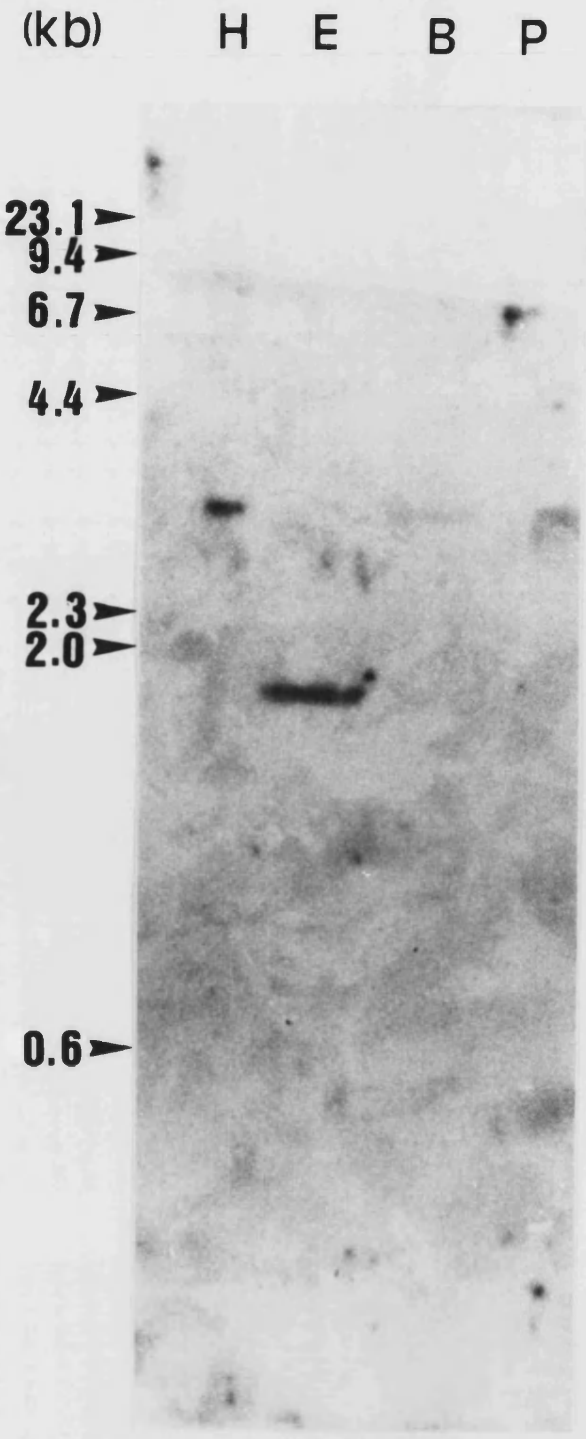

(kb)
H
E
B
P
23.1
9.4
6.7
4.4
2.3
2.0
0.6

Figure 3.6: Hybridisation of double enzyme digests of Tp.acidophilum DNA with labelled CS-17-mer oligonucleotide

The gel used for the blot shown here is described in Fig. 3.4. Experimental details are given in Section 3.2.3. Arrows in the left-hand margin indicate the migration points of Lambda DNA-HindIII fragments. The tracks are labelled according to the enzymes used for digestion of the Tp.acidophilum DNA.

B: BamHI; E: EcoRI; H: HindIII; P: PstI.

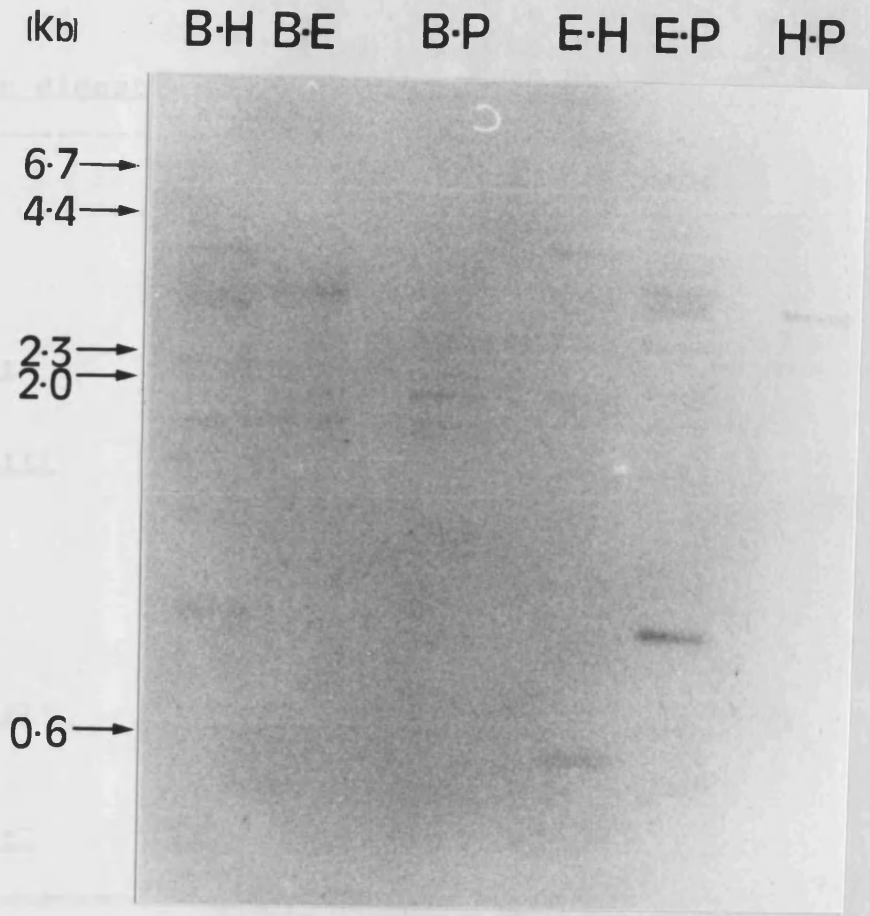

(Kb)
B·H
B·E
B·P
E·H
E·P
H·P
6·7
4·4
2·3
2·0
0·6

Table 3.2: DNA fragments identified by the CS-17-mer and CS-48-mer oligonucleotide probes

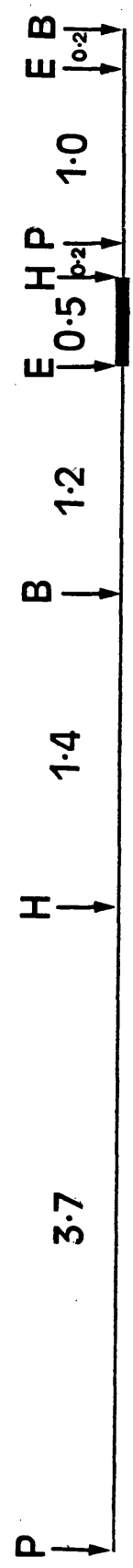
See Sections 3.2.3 and 3.2.5 for experimental details. Where more than one fragment was identified, an asterisk indicates the fragment which fits with the restriction map proposed for the hybridised region (see Figures 3.7 and 3.11).

Restriction digest	Fragments (kb) identified by CS-17-mer	CS-48-mer
BamHI	3.1	1.7
EcoRI	1.7	12.1
HindIII	3.1	4.0
PstI	7.0	3.3
BamHI-EcoRI	3.2 1.7*	1.7
BamHI-HindIII	3.1 2.3 1.7* 0.8	1.7
BamHI-PstI	1.9	1.6
EcoRI-HindIII	0.5	4.0
EcoRI-PstI	0.7	3.3
HindIII-PstI	3.1	3.3

Figure 3.7: Restriction map of region containing the binding site of the CS-17-mer oligonucleotide

The region marked by the dark bar was cloned into Bluescript (see Section 3.2.4. for details). The numbers refer to kilobase pairs between the various restriction sites.

B: BamHI; E: EcoRI; H: HindIII; P: PstI.


E
B
0·2
1·0
H
P
0·2
0·5
E
1·2
B
1·4
H
3·7
P

at RT, once at 37°C and once at 42°C. Four colonies gave strong positive signals; two of these were picked and plasmid prepared from them by the "maxiprep" method outlined in Chapter 2. Initial sequencing of these plasmids, using the forward sequencing primer [5'-GTTTTCCCAGTCACGAC-3'], indicated that both contained the same insert. The complete sequence of the EcoRI-HindIII insert of one of these two plasmids (designated p6H) was then obtained by sequencing with the forward (as above) and reverse [5'-AACAGCTATGACCATG-3'] sequencing primers. This sequence is presented in Fig. 3.8.
No region of the EcoR1-HindIII insert corresponds to the DNA sequence predicted by the N-terminal amino acid sequence of Tp.acidophilum citrate synthase. However, an open reading frame followed by a region rich in pyrimidines (a characteristic of archaebacterial transcription termination [Brown et al. (1989)]) has been identified and is shown in Fig. 3.8. A putative hybridisation site for the CS-17-mer has also been identified (Fig. 3.8).

3.2.5. Use of the CS-48-mer oligonucleotide

The two Southern blots described above were stripped of the CS-17-mer by incubation in 0.4 M NaOH at 42°C for 30 min, and neutralised by incubation in a solution containing 0.1x SSC, 0.1% (w/v) SDS and 0.2 M Tris-HCl (pH 7.5). The blots were hybridised with [γ-^{32}P] end-labelled CS-48-mer at 55°C and washed in 6x SSC containing 1% (w/v) SDS: twice at RT and then twice at 37°C. The autoradiograph obtained for the blot of single enzyme digests is represented in Fig. 3.9 and the autoradiograph obtained for the blot of double enzyme digests is represented in Fig. 3.10.

The DNA fragments, to which the CS-48-mer hybridised, are listed in Table 3.2 and a restriction map of this region presented in Fig. 3.11.

Fig 3.8: The DNA sequence of the EcoRI-HindIII insert of plasmid p6H

An open reading frame has been identified. Underlined nucleotides may play a role in transcription termination. The region of nucleotides highlighted with asterisks and bars may act as a hybridisation site for the CS-17-mer (an asterisk indicates a direct match and a bar indicates a mismatch).

```
 K  L  R  A  T  S  I  S  T  I  K  I  V  T  R  G  H  D  R  G
AAGCTTCGGGCCACGTCAATATCCACCATAAAAATCGTAACGCGGGGACATGATAGAGGA
HindIII          20                   40                   60

 G  G  R  T  R  L  R  K  P  H  S  I  E  I  L  S  V  L  A  V
GGCGGGAGAACCAGATTGAGGAAGCCGCACTCCATTGAAATACTATCAGTTCTGGCGGTA
                 80                  100                  120

 L  S  V  L  M  F  L  S  S  A  S  S  I  I  Y  L  L  L  F  L
TTATCGGTGCTGATGTTTTTATCATCCGCCAGTTCCATAATATATCTTCTGTTATTTTTA
                140                  160                  180

 S  H  R stop
TCGCATCGCTGAtcatccttgttttctccaagccgaacgcttctttgacaatgcggtgcc
                200                  220                  240

           ****-*-**--**-***
atgttccagttccaggggcatcggggatatgtatgccggaaccatattcgttggatttgc
                260                  280                  300

ctccgttctggatgagatagtcgctttccgtcatcttcgcttcatattccgtcacccgga
                320                  340                  360

aataggcataggtcgtctgtggagggttcgaatctcataacgatggcattcttcacggtt
                380                  400                  420

atcgtcgcggtaagctatctggaagttgccggagaattc
                440                EcorI
```

Figure 3.9: Hybridisation of single enzyme digests of Tp.acidophilum DNA with labelled CS-48-mer oligonucleotide

The gel used for the blot shown here is described in Fig. 3.3. Experimental details are given in Section 3.2.5. Arrows in the left-hand margin indicate the migration points of Lambda DNA-HindIII fragments. The tracks are labelled according to the enzyme used for digestion of the Tp.acidophilum DNA.

B: BamHI; E: EcoRI; H: HindIII; P: PstI.

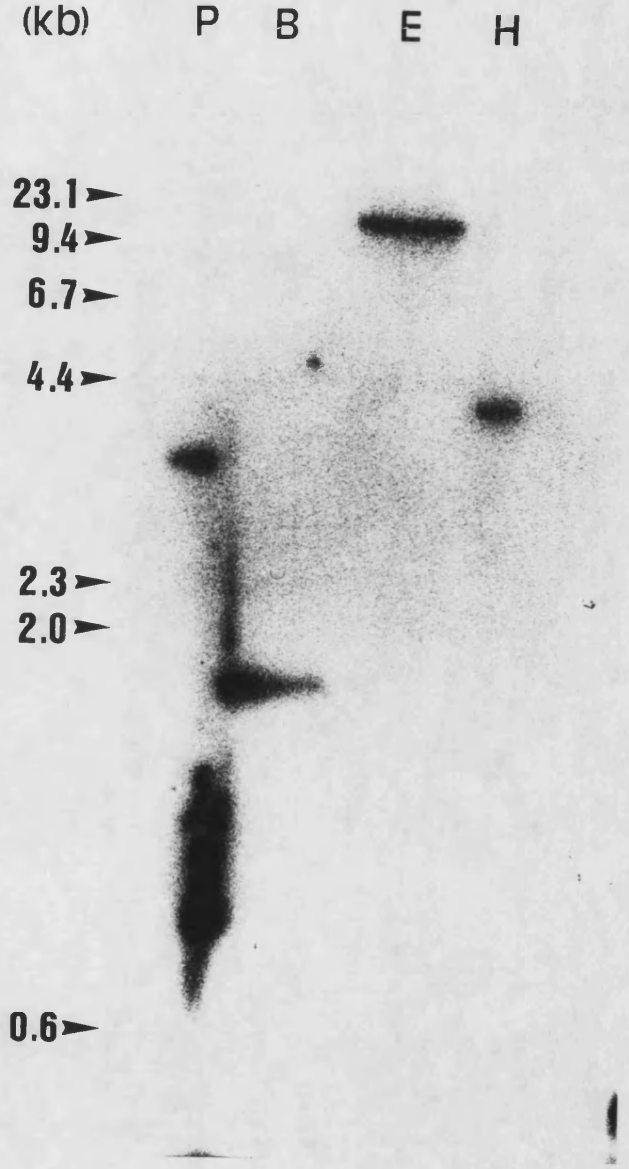

(kb)
P
B
E
H
23.1
9.4
6.7
4.4
2.3
2.0
0.6

Figure 3.10: Hybridisation of double enzyme digests of Tp.acidophilum DNA with labelled CS-48-mer oligonucleotide

The gel used for the blot shown here is described in Fig. 3.4. Experimental details are given in Section 3.2.5. Arrows in the left-hand margin indicate the migration points of Lambda DNA-HindIII fragments, and arrows in the right-hand margin indicate the positions of bands detected by the CS-48-mer in single enzyme digests (Fig. 3.9). The tracks are labelled according to the enzymes used for digestion of the Tp.acidophilum DNA.

B: BamHI; E: EcoRI; H: HindIII; P: PstI.

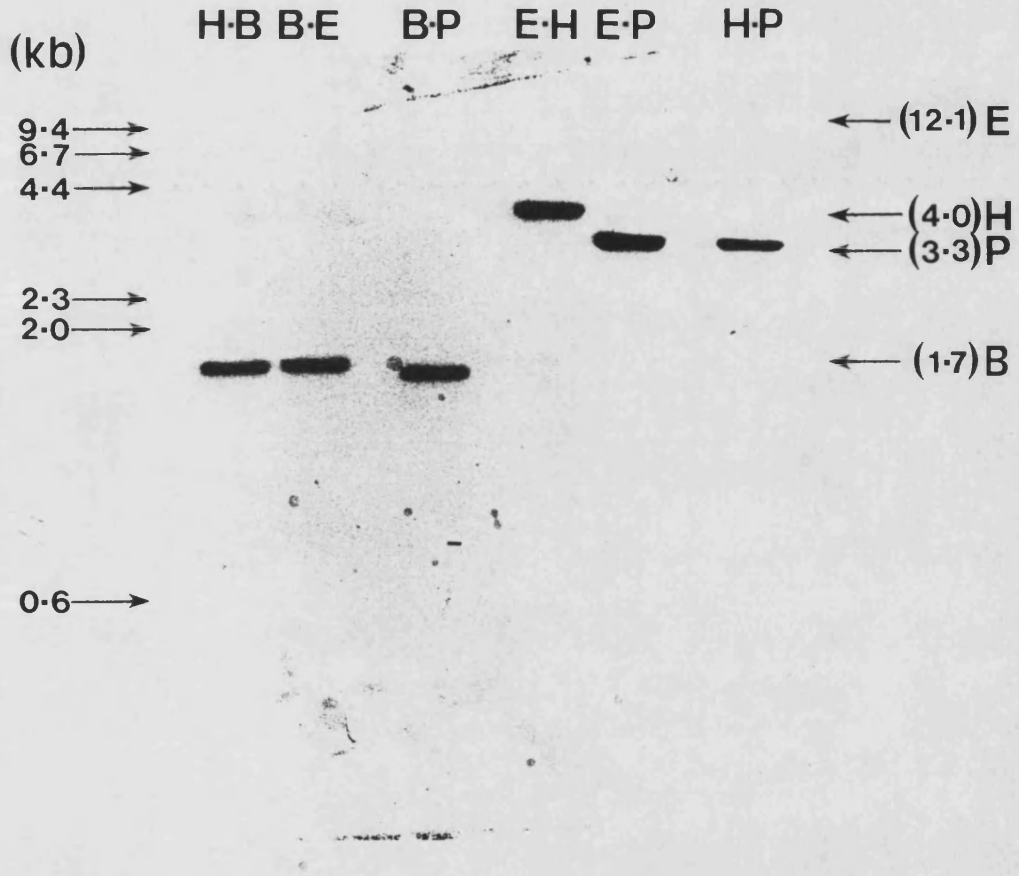
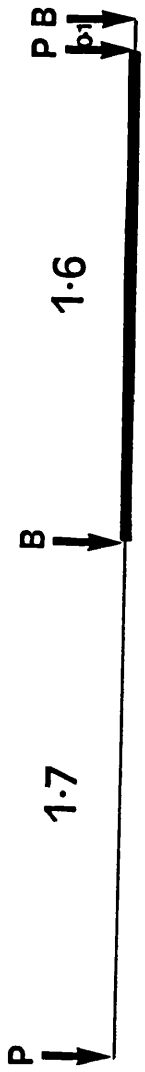

H·B B·E B·P E·H E·P H·P
(kb)
9·4
6·7
4·4
2·3
2·0
0·6
(12·1) E
(4·0) H
(3·3) P
(1·7) B

Figure 3.11: Restriction map of region containing the binding site of the CS-48-mer oligonucleotide

The region marked by the dark bar was cloned into pUC19 (see Section 3.2.6 for details). The numbers refer to kilobase pairs between the various restriction sites.

B: BamHI; P: PstI.


P B
1·6
B
1·7
P

3.2.6. Cloning and sequencing of the 1.6-kb BamHI-PstI fragment identified by the CS-48-mer oligonucleotide

As the BamHI-PstI fragment (1.6-kb), identified by the CS-48-mer, was smaller than the fragments produced with either BamHI (1.65-kb) and PstI (6.7-kb) alone, it was clear that this fragment had two different ends (i.e. a BamHI and a PstI end). The other five combinations of enzymes gave fragments equal in size to those obtained by digestion with just one of the two enzymes involved (Table 3.2). Therefore, since the 1.6-kb BamHI-PstI fragment was the only fragment compatible with the directional cloning strategy to be employed, it was selected for cloning.

Approximately 25 μg Tp.acidophilum genomic DNA was digested with BamHI and PstI, and separated on a 0.8% (w/v) LMP agarose gel. DNA of fragment-size approximately 1.6-kb was recovered by the freeze-squeeze method and purified on a NENsorb column. This size-selected DNA was then ligated into BamHI-PstI-cleaved pUC19 vector and transformed into competent E.coli DH5α cells. Approximately 1 000 transformants were transferred on to nylon and hybridised with [γ-^{32}P] end-labelled CS-48-mer. The filters were washed in 6x SSC containing 1% (w/v) SDS: once at RT, once at 37°C and then once at 45°C. Ten colonies gave strong positive signals; four of these were picked and plasmid was recovered from them by the "maxiprep" method outlined in Chapter 2. Sequencing of all four clones, using the CS-17-mer as a primer, confirmed the presence of sequence corresponding to that predicted by the N-terminal amino acid sequence. This indicated that all four clones contained the 5′ end of the Tp.acidophilum citrate synthase gene. Further sequencing of one clone (designated pTaCS1) demonstrated that it encoded only the first 212 amino acids of citrate synthase. The gene sequence then terminated at the PstI site of pTaCS1 (Fig. 3.12).

Figure 3.12: DNA sequence of the 5′ end of the Tp.acidophilum citrate synthase gene carried on plasmid pTaCS1

The dots highlight the N-terminal residues of the Tp.acidophilum citrate synthase [Smith et al. (1987)] used as a basis for the design of the CS-48-mer.

```
TATTCTCCGCACGGTACGGACTTCGCTATTCTTGCCGATTGGGTCTAAATAGTATATTTT
          -150                  -130                  -110

TAATCTATTTATTTGAATATTAAAATATGTATACCGATCCCAAAATTATACGACGAAATT
          -90                   -70                   -50

                                                  .  .  .  .  .
                                                  P  E  T  E  E    5
TCAAATAGCGTTAAAAGATATCATATATGTAGAGGTGTATTAATGCCAGAAACTGAAGAA
          -30                   -10                   10

 .  .  .  .  .  .  .  .  .  .  .  .
 I  S  K  G  L  E  D  V  N  I  K  W  T  R  L  T  T  I  D  G   25
ATTAGCAAAGGGCTTGAGGATGTCAATATAAAGTGGACAAGGCTCACGACCATCGACGGA
          30                    50                    70

 N  K  G  I  L  R  Y  G  G  Y  S  V  E  D  I  I  A  S  G  A   45
AATAAGGGAATCCTGCGATACGGGGGCTATTCTGTCGAAGATATAATTGCCTCTGGGGCA
          90                    110                   130

 Q  D  E  E  I  Q  Y  L  F  L  Y  G  N  L  P  T  E  Q  E  L   65
CAGGATGAGGAGATCCAGTATCTGTTCCTGTATGGGAACCTTCCAACTGAACAGGAGCTC
          150                   170                   190   SacI

 R  K  Y  K  E  T  V  Q  K  G  Y  K  I  P  D  F  V  I  N  A   85
AGAAAATACAAGGAAACCGTCCAGAAGGGTTATAAGATACCAGATTTTGTCATAAACGCC
          210                   230                   250

 I  R  Q  L  P  R  E  S  D  A  V  A  M  Q  M  A  A  V  A  A  105
ATAAGGCAGCTGCCAAGGGAATCGGATGCTGTGGCAATGCAGATGGCCGCGGTTGCAGCC
          270                   290                   310

 M  A  A  S  E  T  K  F  K  W  N  K  D  T  D  R  D  V  A  A  125
ATGGCAGCATCGGAGACAAAGTTCAAGTGGAACAAGGATACGGACAGAGACGTTGCAGCA
          330                   350                   370

 E  M  I  G  R  M  S  A  I  T  V  N  V  Y  R  H  I  M  N  M  145
GAAATGATCGGCAGGATGTCAGCGATAACCGTGAACGTGTACAGGCACATAATGAACATG
          390                   410                   430

 P  A  E  L  P  K  P  S  D  S  Y  A  E  S  F  L  N  A  A  F  165
CCGGCCGAGCTTCCGAAACCATCAGACAGCTACGCTGAGAGCTTCCTGAACGCTGCGTTT
          450                   470                   490

 G  R  K  A  T  K  E  E  I  D  A  M  N  T  A  L  I  L  Y  T  185
GGAAGGAAGGCAACAAAGGAAGAGATAGACGCCATGAATACGGCACTGATTCTCTACACA
          510                   530                   550

 D  H  E  V  P  A  S  T  T  A  G  L  V  A  V  S  T  L  S  D  205
GACCATGAGGTGCCTGCATCTACCACGGCAGGCTTGGTCGCGGTATCAACCCTCTCAGAC
          570                   590                   610

 M  Y  S  G  I  T  A                                        212
ATGTACTCCGGCATAACTGCAG
          630     PstI
```

3.2.7. Identifying and cloning the 3′ portion of the Tp.acidophilum citrate synthase gene

Sequencing of pTaCS1 indicated the presence of a SacI site 190-bp downstream from the 5′ end of the citrate synthase gene. A 16-mer oligonucleotide [AGA AGG GTT ATA AGA T] (designated CS-16-mer) corresponding to a region of sequence between the known SacI and PstI sites was designed to probe for the 3′ end of the gene (see Fig. 3.15). Tp.acidophilum genomic DNA (2.5 μg) was digested with SacI and KpnI and separated on a 0.5% (w/v) agarose gel (Fig. 3.13) (other digests on this gel were included as controls). The gel was then blotted onto GeneScreen Plus, and the membrane hybridised with [γ-^{32}P] end-labelled CS-16-mer overnight at 37°C. It was washed twice in 6x SSC containing 1% (w/v) SDS at RT. The resulting autoradiograph is represented in Fig. 3.14.

The fragments identified by the CS-16-mer are presented in Table 3.3. From these results, it appeared that the 3′ end of the Tp.acidophilum gene was carried on a 2.7-kb SacI-KpnI fragment, and so this fragment was selected for cloning.

Approximately 25 μg Tp.acidophilum genomic DNA was digested with SacI and KpnI and separated on a 0.8% (w/v) LMP agarose gel. DNA of fragment-size approximately 2.7-kb was recovered by the freeze-squeeze method and purified on a NENsorb column. This DNA was then ligated into SacI-KpnI-cleaved Bluescript vector and transformed into competent E.coli TG1 cells. Approximately 100 colonies were transferred to nylon and hybridised with [γ-^{32}P] end-labelled CS-16-mer. A strong positive signal was obtained from five clones. Partial sequencing of plasmid (designated pTaCS2) prepared from one of these clones confirmed that it contained the portion of the Tp.acidophilum citrate synthase gene, downstream of the known Sac1 site (Fig. 3.15).

Figure 3.13: Digests of Tp.acidophilum genomic DNA separated on a 0.6% (w/v) agarose gel

The sizes of the Lambda DNA-HindIII fragments are indicated in the left-hand margin.

E: EcoRI; H: HindIII; S: SacI; K: KpnI.

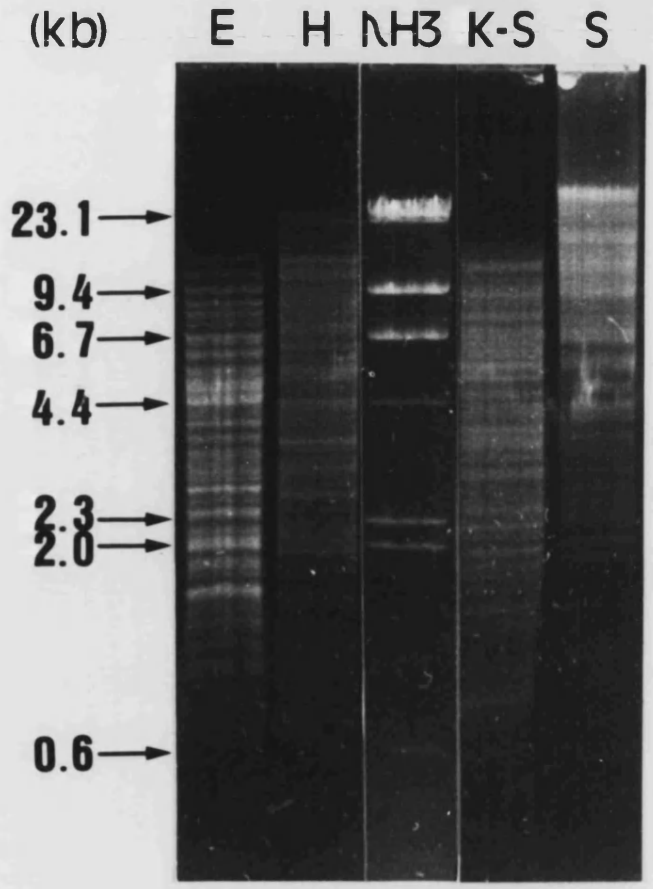

(kb)
E
H
ΛH3
K-S
S
23.1
9.4
6.7
4.4
2.3
2.0
0.6

Figure 3.14: Hybridisation of Tp.acidophilum DNA with labelled CS-16-mer oligonucleotide

The gel used for the blot shown here is presented and described in Fig. 3.13. Experimental details are given in the Section 3.2.7. Arrows in the left-hand margin indicate the migration points of Lambda DNA-HindIII fragments. Tracks are labelled according to the enzyme(s) used for digestion of the Tp.acidophilum DNA.

E: EcoRI; H: HindIII; S: SacI; K: KpnI.

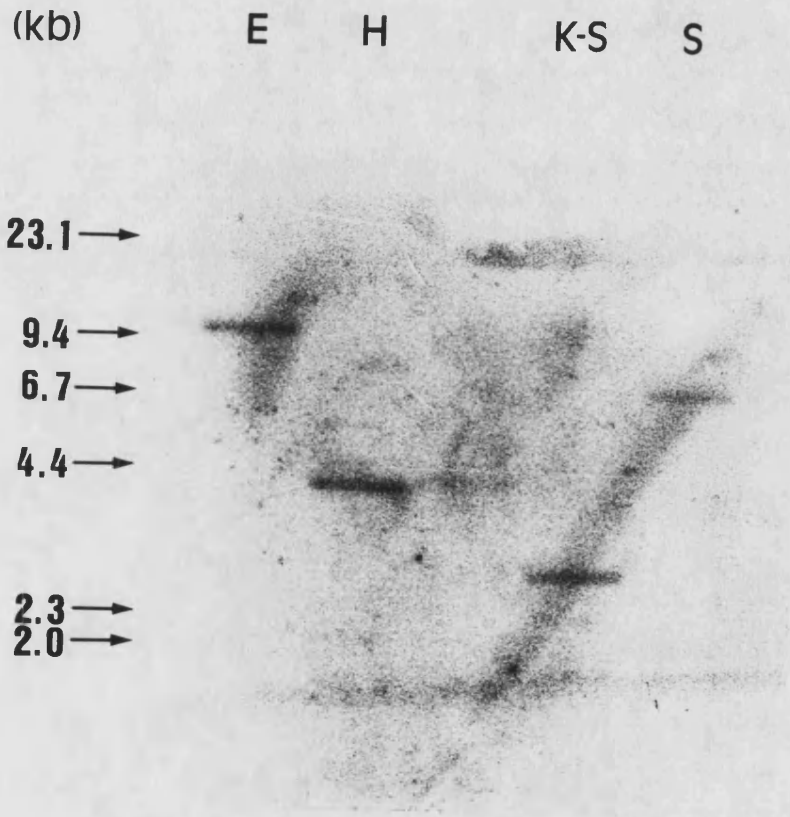

(kb)
E
H
K-S
S
23.1
9.4
6.7
4.4
2.3
2.0

Table 3.3: DNA fragments identified by the CS-16-mer oligonucleotide (see Fig. 3.14).

Experimental details are given in Section 3.2.7.

Restriction digest	Fragments size (kb)
EcoRI	10.5
SacI	6.7
HindIII	4.0
SacI-KpnI	2.7

Figure 3.15: The DNA sequence of the 3′ end of the Tp.acidophilum citrate synthase gene carried on plasmid pTaCS2

The sequence of nucleotides to which the CS-16-mer oligonucleotide was designed is shown in lower case. The asterisk denotes a stop codon.

```
 E  L  R  K  Y  K  E  T  V  Q  K  G  Y  K  I  P  D  F  V  I   83
GAGCTCAGAAAATACAAGGAAACCGTCCagaagggttataagatACCAGATTTTGTCATA
  SacI          210                  230                  250
 N  A  I  R  Q  L  P  R  E  S  D  A  V  A  M  Q  M  A  A  V  103
AACGCCATAAGGCAGCTGCCAAGGGAATCGGATGCTGTGGCAATGCAGATGGCCGCGGTT
                270                  290                  310
 A  A  M  A  A  S  E  T  K  F  K  W  N  K  D  T  D  R  D  V  123
GCAGCCATGGCAGCATCGGAGACAAAGTTCAAGTGGAACAAGGATACGGACAGAGACGTT
                330                  350                  370
 A  A  E  M  I  G  R  M  S  A  I  T  V  N  V  Y  R  H  I  M  143
GCAGCAGAAATGATCGGCAGGATGTCAGCGATAACCGTGAACGTGTACAGGCACATAATG
                390                  410                  430
 N  M  P  A  E  L  P  K  P  S  D  S  Y  A  E  S  F  L  N  A  163
AACATGCCGGCCGAGCTTCCGAAACCATCAGACAGCTACGCTGAGAGCTTCCTGAACGCT
                450                  470                  490
 A  F  G  R  K  A  T  K  E  E  I  D  A  M  N  T  A  L  I  L  183
GCGTTTGGAAGGAAGGCAACAAAGGAAGAGATAGACGCCATGAATACGGCACTGATTCTC
                510                  530                  550
 Y  T  D  H  E  V  P  A  S  T  T  A  G  L  V  A  V  S  T  L  203
TACACAGACCATGAGGTGCCTGCATCTACCACGGCAGGCTTGGTCGCGGTATCAACCCTC
                570                  590                  610
 S  D  M  Y  S  G  I  T  A  A  L  A  A  L  K  G  P  L  H  G  223
TCAGACATGTACTCCGGCATAACTGCAGCTCTTGCGGCCCTCAAGGGTCCGCTGCATGGC
                630                  650                  670
 G  A  A  E  A  A  I  A  Q  F  D  E  I  K  D  P  A  M  V  E  243
GGTGCAGCTGAGGCAGCAATAGCGCAGTTCGATGAGATAAAGGATCCCGCCATGGTGGAG
                690                  710                  730
 K  W  F  N  D  N  I  I  N  G  K  K  R  L  M  G  F  G  H  R  263
AAGTGGTTCAACGACAACATAATAAACGGAAAGAAGAGACTCATGGGCTTCGGCCACAGG
                750                  770                  790
 V  Y  K  T  Y  D  P  R  A  K  I  F  K  G  I  A  E  K  L  S  283
GTATACAAGACCTACGATCCAAGGGCAAAGATATTCAAGGGCATCGCTGAGAAACTCTCC
                810                  830                  850
 S  K  K  P  E  V  H  K  V  Y  E  I  A  T  K  L  E  D  F  G  303
AGCAAGAAGCCAGAGGTGCACAAGGTATATGAGATCGCCACGAAGCTTGAAGACTTTGGC
                870                  890                  910
 I  K  A  F  G  S  K  G  I  Y  P  N  T  D  Y  F  S  G  I  V  323
ATCAAGGCATTCGGTTCAAAGGGCATATATCCGAACACGGATTACTTCTCCGGAATAGTA
                930                  950                  970
 Y  M  S  I  G  F  P  L  R  N  N  I  Y  T  A  L  F  A  L  S  343
TACATGTCCATAGGATTCCCGCTGAGGAACAACATATACACGGCGCTCTTCGCGCTATCC
                990                 1010                 1030
 R  V  T  G  W  Q  A  H  F  I  E  Y  V  E  E  Q  Q  R  L  I  363
AGGGTTACAGGGTGGCAGGCACACTTCATAGAGTACGTGGAGGAACAGCAGAGACTGATT
               1050                 1070                 1090
 R  P  R  A  V  Y  V  G  P  A  E  R  K  Y  V  P  I  A  E  R  383
AGGCCGAGAGCCGTTTACGTCGGTCCTGCAGAGAGAAAGTACGTACCCATCGCTGAAAGA
               1110                 1130                 1150
 K  *                                                        384
AAGTGAAAACACTTTTCTCAAATTTTTATTAATATATTTTATATACAACGGCGCTTGTTG
               1170                 1190                 1210

TGGAAATTTCGATTAGCGATGGTCTGTGTTTTCGGATCC
               1230                 BamHI
```

3.2.8. Construction of plasmid pTaCS19

A construct (designated pTaCS19) containing the complete Tp.acidophilum citrate synthase gene was assembled by the scheme described below and depicted in Fig. 3.16.

pTaCS1 was digested with BamHI and SacI and the digest separated on a 0.8% (w/v) LMP agarose gel. The 1.2-kb BamHI-SacI fragment was recovered by the freeze-squeeze method and cloned into pUC19 to generate the clone pTaCS3. pTaCS2 was digested with SacI and then partially with BamHI, and the digest separated on a 0.8% (w/v) LMP agarose gel. The 1.2-kb BamHI-SacI fragment was recovered by the freeze-squeeze method and cloned into Bluescript to generate the clone pTaCS4. pTaCS4 was digested with SacI and EcoRI and the digest separated on a 0.8% (w/v) LMP agarose gel. The 1.2-kb SacI-EcoRI fragment was recovered by the freeze-squeeze method and cloned into SacI-EcoR1-digested pTaCS3 to yield a construct, designated pTaCS19. Sequencing of pTaCS19 (see below) confirmed that it contained the complete citrate synthase gene, and that no rearrangements had occurred.

3.2.9. Sequencing of the Tp.acidophilum citrate synthase gene

The gene was subcloned from pTaCS19 into M13 and the entire coding region sequenced on both DNA strands, using the strategy depicted in Fig. 3.17. A restriction map of the Tp.acidophilum citrate synthase gene and the immediate flanking regions is presented in Fig. 3.17.

The complete DNA sequence of the Tp.acidophilum gene, its immediate 5′- and 3′-flanking regions, and the derived amino acid sequence of the protein are presented in Fig. 3.18. This protein, with M_r 42 942, corresponds to the size determined by Smith et al. (1987) for citrate synthase purified from Tp.acidophilum [M_r=43 000 (± 2 000)].

Further information obtained from the DNA and

Figure 3.16: Construction of plasmid pTaCS19 containing the complete gene for citrate synthase from Tp.acidophilum

(a) The 1.2-kb BamHI-SacI fragment (shaded) was cloned into pUC19 to generate the clone pTaCS3. (b) The 1.2-kb SacI-partial-BamHI fragment of pTaCS2 (hatched) was subcloned into Bluescript to generate the clone pTaCS4. (c) The 1.2-kb SacI-EcoRI fragment of pTaCS4 was cloned into SacI-EcoRI-digested pTaCS3 to yield pTaCS19.

B: BamHI; E: EcoRI; P: PstI; S: SacI; K: KpnI.
N.B. not to scale.

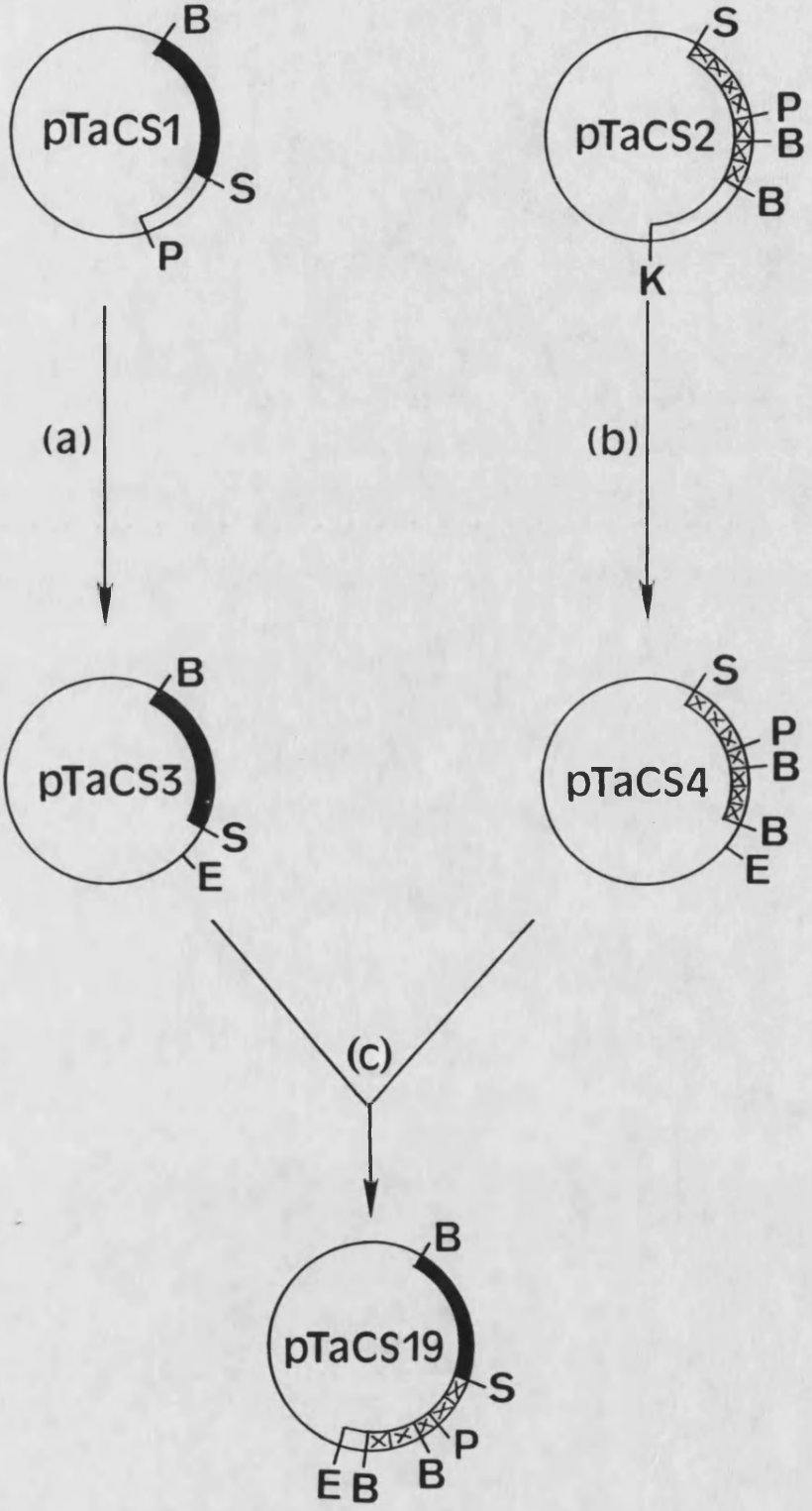

B
pTaCS1
S
P
S
P
B
pTaCS2
B
K
(a)
(b)
B
pTaCS3
S
E
S
P
B
pTaCS4
B
E
(c)
B
pTaCS19
S
P
E
B
B

Figure 3.17: Restriction map of Tp.acidophilum DNA containing the citrate synthase gene

The coding region is marked by the dark bar. Figures refer to the number of base-pairs between the various restriction sites. The horizontal arrows indicate the direction of the individual sequencing runs. Most of the sequence was generated by progressively "walking" along the gene with oligonucleotides (these were designed according to previously derived sequence data). The 532 bp BamHI-BamHI fragment, however, was subcloned and sequenced with the standard M13 forward and reverse sequencing primers.

B: BamHI; E: EcoRV; P: PstI; S: SacI; K: KpnI.

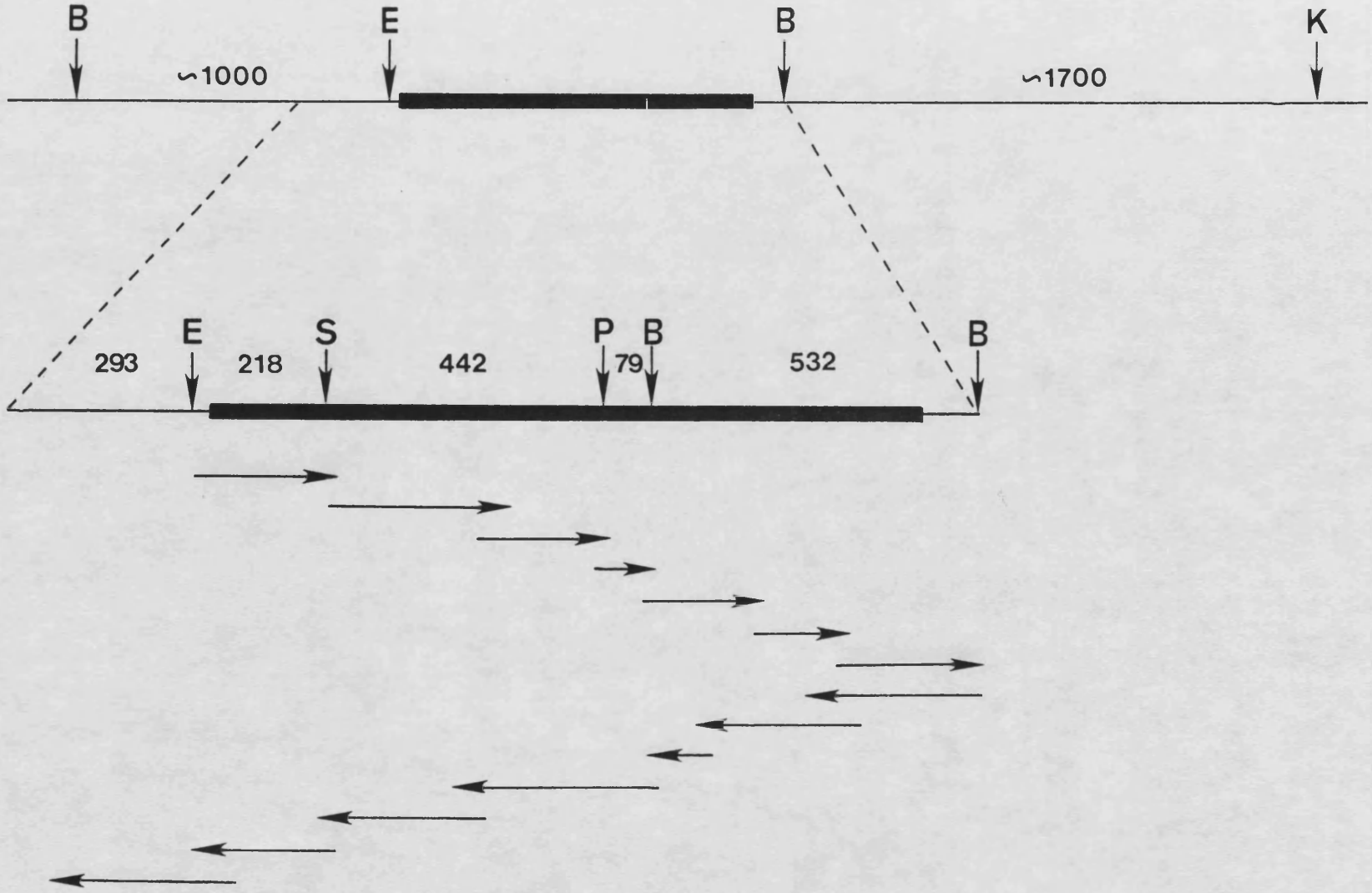

B
E
B
K
~1000
~1700
E
S
P
B
B
293
218
442
79
532

Figure 3.18: The complete DNA sequence of the Tp.acidophilum citrate synthase gene (upper case) and its immediate flanking regions (lower case)

The derived amino acid sequence is shown. The asterisk denotes a stop codon.

cgctccatatcggcgtcgcttcgatatttcatcataactatattaccgcttttgtttctcatcacgttattcctctggattgttctacggctttggcatattaatct
-269 -250 -230 -210 -190 -170

tattctccgcacggatcggacttcgctattcttgccgattgggtctaaatagtatatttttaatctatttatttgaatattaaaatatgtataccgatcccaaaattatacgacgaaatt
-150 -130 -110 -90 -70 -50

P E T E E I S K G L E D V N I K W T R L T T I D G 25
tcaaatagcgttaaaagatatcatatatgtagaggtgtattaATGCCAGAAACTGAAGAAATTAGCAAAGGGCTTGAGGATGTCAATATAAAGTGGACAAGGCTCACGACCATCGACGGA
-30 -10 10 30 50 70

N K G I L R Y G G Y S V E D I I A S G A Q D E E I Q Y L F L Y G N L P T E Q E L 65
AATAAGGGAATCCTGCGATACGGGGGCTATTCTGTCGAAGATATAATTGCCTCTGGGGCACAGGATGAGGAGATCCAGTATCTGTTCCTGTATGGGAACCTTCCAACTGAACAGGAGCTC
90 110 130 150 170 190

R K Y K E T V Q K G Y K I P D F V I N A I R Q L P R E S D A V A M Q M A A V A A 105
AGAAAATACAAGGAAACCGTCCAGAAGGGTTATAAGATACCAGATTTTGTCATAAACGCCATAAGGCAGCTGCCAAGGGAATCGGATGCTGTGGCAATGCAGATGGCCGCGGTTGCAGCC
210 230 250 270 290 310

M A A S E T K F K W N K D T D R D V A A E M I G R M S A I T V N V Y R H I M N M 145
ATGGCAGCATCGGAGACAAAGTTCAAGTGGAACAAGGATACGGACAGAGACGTTGCAGCAGAAATGATCGGCAGGATGTCAGCGATAACCGTGAACGTGTACAGGCACATAATGAACATG
330 350 370 390 410 430

P A E L P K P S D S Y A E S F L N A A F G R K A T K E E I D A M N T A L I L Y T 185
CCGGCCGAGCTTCCGAAACCATCAGACAGCTACGCTGAGAGCTTCCTGAACGCTGCGTTTGGAAGGAAGGCAACAAAGGAAGAGATAGACGCCATGAATACGGCACTGATTCTCTACACA
450 470 490 510 530 550

D H E V P A S T T A G L V A V S T L S D M Y S G I T A A L A A L K G P L H G G A 225
GACCATGAGGTGCCTGCATCTACCACGGCAGGCTTGGTCGCGGTATCAACCCTCTCAGACATGTACTCCGGCATAACTGCAGCTCTTGCGGCCCTCAAGGGTCCGCTGCATGGCGGTGCA
570 590 610 630 650 670

A E A A I A Q F D E I K D P A M V E K W F N D N I I N G K K R L M G F G H R V Y 265
GCTGAGGCAGCAATAGCGCAGTTCGATGAGATAAAGGATCCCGCCATGGTGGAGAAGTGGTTCAACGACAACATAATAAACGGAAAGAAGAGACTCATGGGCTTCGGCCACAGGGTATAC
690 710 730 750 770 790

K T Y D P R A K I F K G I A E K L S S K K P E V H K V Y E I A T K L E D F G I K 305
AAGACCTACGATCCAAGGGCAAAGATATTCAAGGGCATCGCTGAGAAACTCTCCAGCAAGAAGCCAGAGGTGCACAAGGTATATGAGATCGCCACGAAGCTTGAAGACTTTGGCATCAAG
810 830 850 870 890 910

A F G S K G I Y P N T D Y F S G I V Y M S I G F P L R N N I Y T A L F A L S R V 345
GCATTCGGTTCAAAGGGCATATATCCGAACACGGATTACTTCTCCGGAATAGTATACATGTCCATAGGATTCCCGCTGAGGAACAACATATACACGGCGCTCTTCGCGCTATCCAGGGTT
930 950 970 990 1010 1030

T G W Q A H F I E Y V E E Q Q R L I R P R A V Y V G P A E R K Y V P I A E R K * 384
ACAGGGTGGCAGGCACACTTCATAGAGTACGTGGAGGAACAGCAGAGACTGATTAGGCCGAGAGCCGTTTACGTCGGTCCTGCAGAGAGAAAGTACGTACCCATCGCTGAAAGAAAGTGA
1050 1070 1090 1110 1130 1150

aaacacttttctcaaatttttattaatatattttatatacaacggcgcttgttgtggaaatttcgattagcgatggtctgtgttttcggatcc
1170 1190 1210 1230 1250

protein sequences is presented in Chapter 4 and 5, respectively.

3.2.10. Specificity of the CS-17-mer oligonucleotide

An autoradiograph of a Southern blot of digested Tp.acidophilum DNA, probed with the CS-17-mer and washed at 37°C in 6x SSC containing 1% (w/v) SDS, is presented in Fig. 3.19. At this low stringency, in addition to hybridising to fragments listed in column 1 of Table 3.2, the probe hybridised to (PstI, BamHI, and HindIII) fragments that carry the citrate synthase gene (as identified by the CS-48-mer and listed in the second column of Table 3.2). When the filter was washed at 45°C, the probe was removed from the citrate synthase gene, but remained hybridised to the fragments listed in the first column of Table 3.2 (see Section 3.2.3).

3.3. Discussion

Two oligonucleotides, the long non-redundant CS-48-mer and the short redundant CS-17-mer, both based on the N-terminal amino acid sequence of the protein, were used to probe for the 5′ portion of the gene encoding Tp.acidophilum citrate synthase. As can be seen from Table 3.2 these probes hybridise to two separate regions.

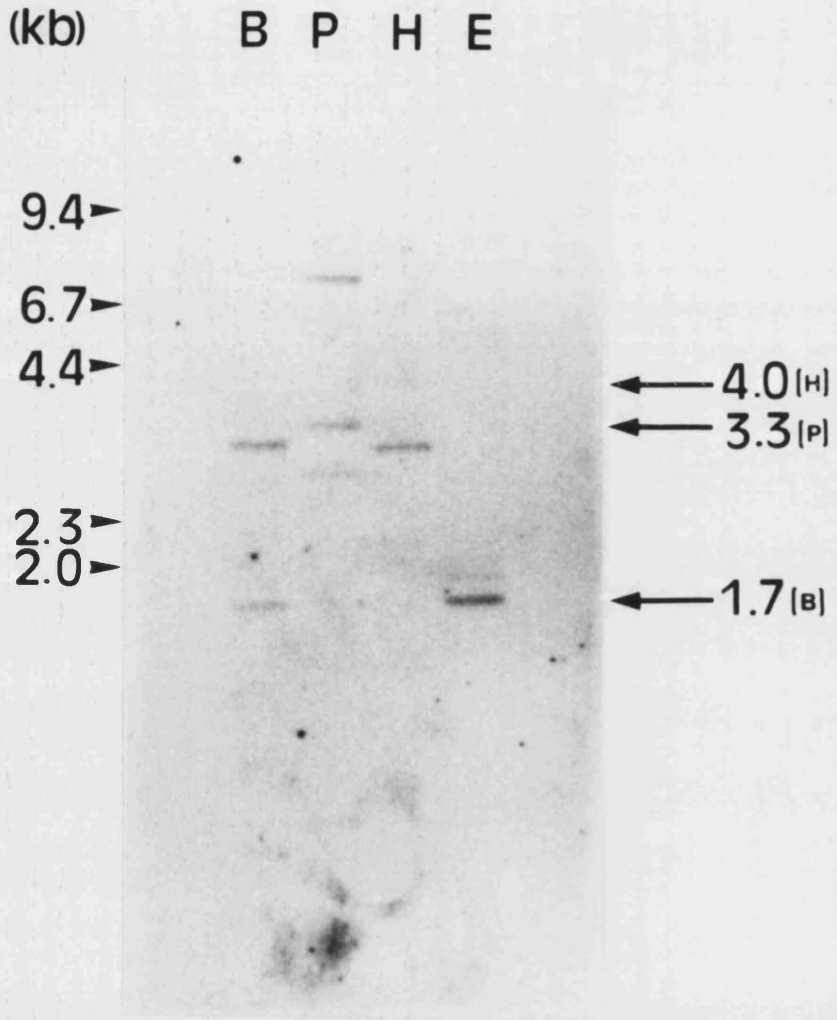
The CS-17mer hybridised only weakly to the blot of double enzyme digests of Tp.acidophilum DNA (Fig. 3.6). However, it did hybridise strongly to the blot of single enzyme digests (Fig. 3.5) and so this poor hybridisation was probably due to experimental conditions (possibly lower concentration of probe) rather than due to the nature of the probe itself.

Sequencing of two fragments, one identified by the CS-17-mer and one identified by the CS-48-mer, indicated that only the latter fragment contained sequence corresponding to the N-terminus of the Tp.acidophilum citrate synthase. So, whereas the CS-48-mer was specific for the Tp.acidophilum citrate synthase gene, the

Figure 3.19: Hybridisation of Tp.acidophilum DNA with labelled CS-17-mer oligonucleotide

A Southern blot of digested Tp acidophilum DNA after hybridisation with the CS-17-mer and washing at low stringency (experimental details are given in Section 3.2.10). Arrows in the left-hand margin indicate the migration points of Lambda DNA-HindIII fragments. Arrows in the right-hand margin indicate bands carrying the citrate synthase gene (i.e. those identified by the 48-mer).

B: BamHI; E: EcoRI; H: HindIII; P: PstI.


(kb)
B P H E
9.4
6.7
4.4
2.3
2.0
4.0 [H]
3.3 [P]
1.7 [B]

CS-17-mer only bound to the citrate synthase at low stringency (see Section 3.2.10).

A putative binding site (Fig. 3.8) for one of the oligonucleotides [CCA GAG ACA GAG GAG AT], within the CS-17-mer mixed probe, was observed in the 0.5-kb EcoRI-HindIII fragment. This oligonucleotide has 12 direct matches to the proposed binding site giving an estimated T_m of 40°C (plus additional contributions due to the interactions of the 5 mismatched bases). The highest T_m estimated for the CS-17-mer oligonucleotide binding to the citrate synthase gene [CCA GAA ACT GAA GAA AT] is 46°C, and so could conceivably be lower than the T_m quoted above. These results indicate that a short redundant probe should only be used in conjunction with a second similar probe, designed to identify an independent region of the desired gene [as in Bright et al. (1990)]. Fragments carrying the gene may then be identified as those binding both probes at low stringency.

A comparison of the CS-48-mer sequence with that determined for the 5′ end of the citrate synthase gene indicated that only 8 mismatches (out of the 20 possible mismatch positions) due to incorrect guessing are present in the probe. Thus, a "best guess" approach (using the codon usage of a related organism) has proved extremely successful in this study.

The long non-redundant CS-48-mer probe proved very specific for the gene. However, the initial clone isolated as a result of probing with this oligonucleotide carried only a truncated 5′ portion of the gene. This construct was sequenced, and the information obtained used to design a probe (CS-16-mer) for the cloning of the remaining 3′ portion of the gene. pTaCS19, a construct carrying the complete gene for Tp.acidophilum citrate synthase, was then assembled by joining the two separate portions present in pTaCS1 and pTaCS2 at a common SacI site. Sequencing of pTaCS19 confirmed that the gene had suffered no rearrangements during this manipulation. Sequence information obtained from pTaCS19 will be discussed further in Chapters 4 and 5.

CHAPTER 4 : ANALYSIS OF THE Tp.acidophilum CITRATE SYNTHASE GENE SEQUENCE

4.1. Introduction

This chapter describes features of the Tp.acidophilum citrate synthase gene and its flanking regions.

4.1.1. Regulatory features of archaebacterial genes

Promoter regions of genes have been defined primarily by the determination of transcription initiation sites and identification of adjacent conserved sequences. The archaebacterial promoter has been shown to consist of a box A sequence and a box B sequence approximately 25 nucleotides apart; transcription initiation occurs within box B, generally at a "G" residue [Zillig et al. (1988); Thomm & Wich (1988); Brown et al. (1989)]. The consensus sequences for box A and for box B appear to be "TTTAT/AA" and "A/TTGA/C", respectively [Thomm & Wich (1988); Zillig et al. (1988)]. Sequences for promoters of the three Tp.acidophilum rRNA genes are also available: the consensus for box A is "CTTATATA", and the box B sequences are "TTGC", "TAC" and "TTCG" [Ree & Zimmerman (1990)].

The format of the archaebacterial promoter resembles the eukaryotic RNA polymerase II promoter, which is characterised by an upstream "TATA" box and a short conserved sequence close to the transcription initiation site [Corden et al. (1980)]. The distances between boxes A and B are approximately the same for the archaebacterial and eukaryotic promoters. In contrast, eubacterial promoters consist of two conserved sequences 10 and 35 bases upstream of the transcription initiation site with no conserved sequence at the initiation site itself [Pribnow (1975)].

Gene transcription termination regions have been defined primarily by determining the 3′ termini of mRNA

transcripts and identifying adjacent conserved sequences. Archaebacterial transcription appears to terminate downstream of regions which have the potential (as transcripts) to form hairpin-loop structures, and which are followed by stretches of thymidines [Brown et al. (1989)].

The archaebacterial ribosome-binding site is believed to consist of a nucleotide sequence that is complementary to the 3'end of the 16S rRNA [reviewed by Brown et al. (1989)]. This ribosome-binding site resembles that found in the eubacteria [Shine & Dalgarno (1974)]. However, whereas in eubacteria the sequence has always been found upstream of the translation initiation codon, in some archaebacterial genes it has been found overlapping the initiation codon or even a short distance downstream of it [Brown et al. (1989) and references therein].

An "ATG" triplet is most frequently used as the translation initiation codon for archaebacterial genes. However, "GTG" and "TGT" (the latter to a lesser extent) are sometimes used [Bokranz & Klein (1987); Cubellis et al. (1989)].

4.1.2. Codon usages of archaebacterial genes

All archaebacterial genes analysed to date use the standard genetic code [Brown et al. (1989)].

The choice of codons in archaebacterial genes, like that in eubacteria and eukaryotes, is dependent on the relative abundance of iso-accepting tRNA species. This means that the codon preference of organisms is highly selective, and so cannot be used to measure long-term evolutionary events (that is across kingdoms) [Woese (1987)]. However, it can measure close evolutionary events, individual organisms and closely-related organisms generally showing similar codon preferences.

4.2. Results

4.2.1. Regulatory features of the Tp.acidophilum citrate synthase gene

Sequencing of pTaCS19 (described in Chapter 3) resulted in the identification of an open reading frame which contained the known sixteen amino acids of the Tp.acidophilum citrate synthase N-terminus. This open reading frame is preceded by an "ATG" triplet (Fig. 4.1), which is proposed to act as the initiation codon for translation.

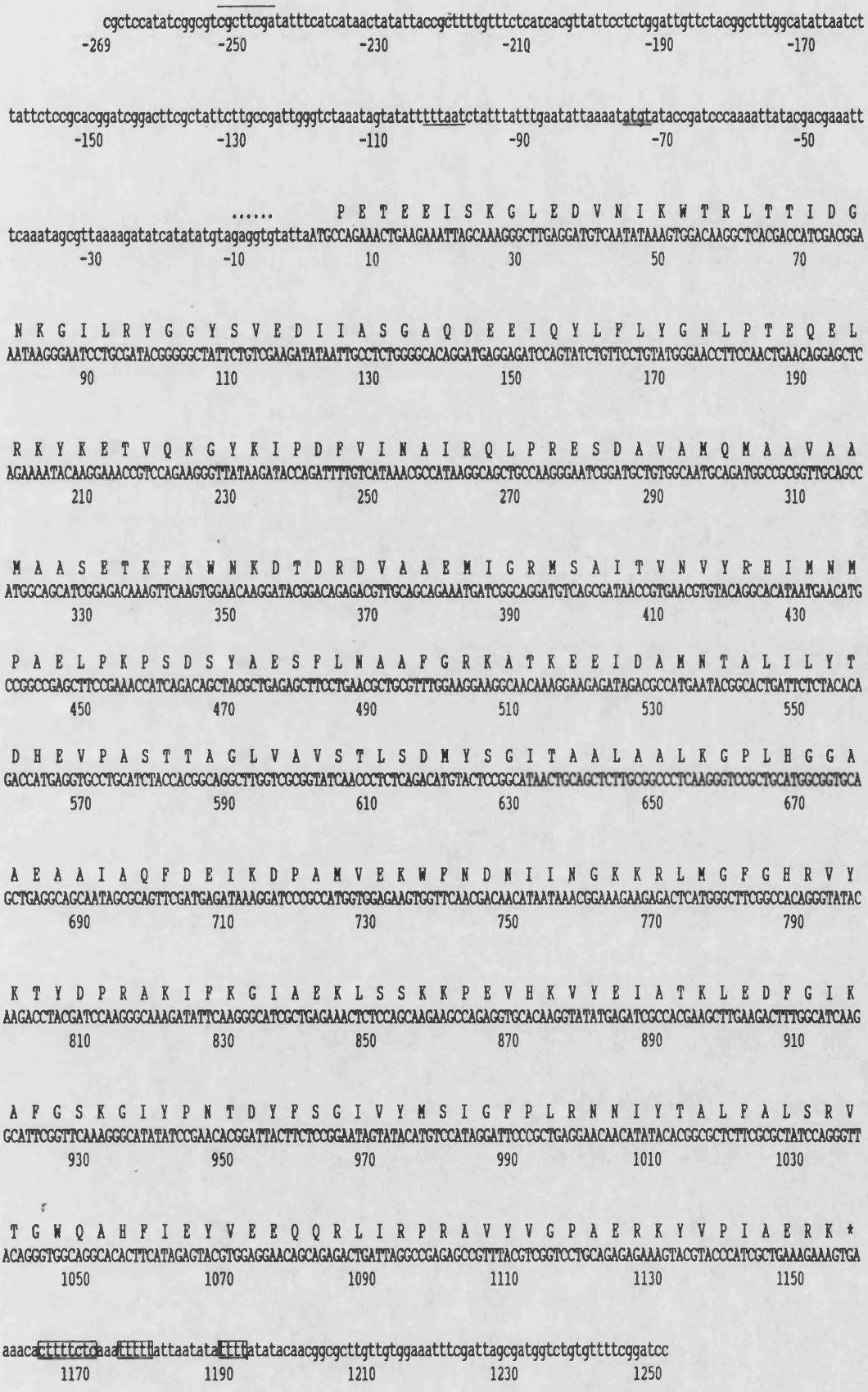
In the absence of the necessary experiments to determine the transcription initiation site of the Tp.acidophilum citrate synthase gene, only tentative conclusions (based on the available archaebacterial consensus sequences) regarding the promoter region can be drawn. The sequence 5'TTTAAT 3' (at position "-104" to "-99") almost matches the consensus for the archaebacterial box A, and is the closest sequence to the Tp.acidophilum box A (see Section 3.1). Looking approximately 25 nucleotides downstream of this putative box A, the sequence corresponding most closely to the consensus for the archaebacterial box B is 5' ATGT 3' (at position "-76" to "-73"). These putative box A and box B sequences are proposed to form the promoter of the Tp.acidophilum gene and are highlighted in Fig. 4.1.

Long stretches of thymidines are found at the 3' end of the Tp.acidophilum citrate synthase gene, immediately downstream of the translation stop codon (TGA) (Fig. 4.1). This region may play a role in transcription termination. Sequences with a propensity to form hair-pin loops have been found in other archaebacterial transcription termination regions, but such sequences are not present at the 3' end of the Tp.acidophilum citrate synthase gene.

The region directly upstream of the start codon of the Tp.acidophilum citrate synthase gene shows

Figure 4.1: The DNA sequence of the Tp.acidophilum citrate synthase gene and the corresponding amino acid sequence

The nucleotides underlined represent putative promoter regions. The nucleotides overlined may play some role in gene expression. The nucleotides boxed may play a role in termination of transcription. The nucleotides overlain with dots represent a potential ribosome-binding site. The asterisk denotes a stop codon.


cgctccatatcggcgtcgcttcgatatttcatcataactatattaccgcttttgtttctcatcacgttattcctctggattgttctacggctttggcatattaatct
-269 -250 -230 -210 -190 -170
tattctccgcacggatcggacttcgctattcttgccgattgggtctaaatagtatatttttaatctatttatttgaatattaaaatatgtataccgatcccaaaattatacgacgaaatt
-150 -130 -110 -90 -70 -50
..... P E T E E I S K G L E D V N I K W T R L T T I D G 25
tcaaatagcgttaaaagatatcatatatgtagaggtgtattaATGCCAGAAACTGAAGAAATTAGCAAAGGGCTTGAGGATGTCAATATAAAGTGGACAAGGCTCACGACCATCGACGGA
-30 -10 10 30 50 70
N K G I L R Y G G Y S V E D I I A S G A Q D E E I Q Y L F L Y G N L P T E Q E L 65
AATAAGGGAATCCTGCGATACGGGGGCTATTCTGTCGAAGATATAATTGCCTCTGGGGCACAGGATGAGGAGATCCAGTATCTGTTCCTGTATGGGAACCTTCCAACTGAACAGGAGCTC
90 110 130 150 170 190
R K Y K E T V Q K G Y K I P D F V I N A I R Q L P R E S D A V A M Q M A A V A A 105
AGAAAATACAAGGAAACCGTCCAGAAGGGTTATAAGATACCAGATTTTGTCATAAACGCCATAAGGCAGCTGCCAAGGGAATCGGATGCTGTGGCAATGCAGATGGCCGCGGTTGCAGCC
210 230 250 270 290 310
M A A S E T K F K W N K D T D R D V A A E M I G R M S A I T V N V Y R H I M N M 145
ATGGCAGCATCGGAGACAAAGTTCAAGTGGAACAAGGATACGGACAGAGACGTTGCAGCAGAAATGATCGGCAGGATGTCAGCGATAACCGTGAACGTGTACAGGCACATAATGAACATG
330 350 370 390 410 430
P A E L P K P S D S Y A E S F L N A A F G R K A T K E E I D A M N T A L I L Y T 185
CCGGCCGAGCTTCCGAAACCATCAGACAGCTACGCTGAGAGCTTCCTGAACGCTGCGTTTGGAAGGAAGGCAACAAAGGAAGAGATAGACGCCATGAATACGGCACTGATTCTCTACACA
450 470 490 510 530 550
D H E V P A S T T A G L V A V S T L S D M Y S G I T A A L A A L K G P L H G G A 225
GACCATGAGGTGCCTGCATCTACCACGGCAGGCTTGGTCGCGGTATCAACCCTCTCAGACATGTACTCCGGCATAACTGCAGCTCTTGCGGCCCTCAAGGGTCCGCTGCATGGCGGTGCA
570 590 610 630 650 670
A E A A I A Q F D E I K D P A M V E K W F N D N I I N G K K R L M G F G H R V Y 265
GCTGAGGCAGCAATAGCGCAGTTCGATGAGATAAAGGATCCCGCCATGGTGGAGAAGTGGTTCAACGACAACATAATAAACGGAAAGAAGAGACTCATGGGCTTCGGCCACAGGGTATAC
690 710 730 750 770 790
K T Y D P R A K I F K G I A E K L S S K K P E V H K V Y E I A T K L E D F G I K 305
AAGACCTACGATCCAAGGGCAAAGATATTCAAGGGCATCGCTGAGAAACTCTCCAGCAAGAAGCCAGAGGTGCACAAGGTATATGAGATCGCCACGAAGCTTGAAGACTTTGGCATCAAG
810 830 850 870 890 910
A F G S K G I Y P N T D Y F S G I V Y M S I G F P L R N N I Y T A L F A L S R V 345
GCATTCGGTTCAAAGGGCATATATCCGAACACGGATTACTTCTCCGGAATAGTATACATGTCCATAGGATTCCCGCTGAGGAACAACATATACACGGCGCTCTTCGCGCTATCCAGGGTT
930 950 970 990 1010 1030
T G W Q A H F I E Y V E E Q Q R L I R P R A V Y V G P A E R K Y V P I A E R K * 384
ACAGGGTGGCAGGCACACTTCATAGAGTACGTGGAGGAACAGCAGAGACTGATTAGGCCGAGAGCCGTTTACGTCGGTCCTGCAGAGAGAAAGTACGTACCCATCGCTGAAAGAAAGTGA
1050 1070 1090 1110 1130 1150
aaacactttttctcaaaatttttattaatatatttttatatacaacggcgcttgttgtggaaatttcgattagcgatggtctgtgttttcggatcc
1170 1190 1210 1230 1250

complementarity to the 3' end of the 16S rRNA (CCUCCACUA) from Tp.acidophilum [Ree et al. (1989)] (see Fig. 4.1). This region is proposed to act as a ribosome-binding site for the Tp.acidophilum citrate synthase gene since it appears to match the format observed for other archaebacterial ribosome-binding sites.

Ree et al. (1989) have suggested that a dodecanucleotide sequence (GTTCCGCTTCGA), found immediately upstream of the promoters for the Tp.acidophilum 16S rRNA gene and preceding open reading frame, may play a role in gene expression. A similar sequence of eight nucleotides (CGCTTCGA) is found upstream of the Tp.acidophilum citrate synthase gene at position "-253" to "-246" and provides support for the proposal of Ree et al. (1989).

4.2.2. Codon usage of Tp.acidophilum

The codon usage of the Tp.acidophilum citrate synthase gene is shown in Table 4.1. For comparison this has been presented along with codon usages of other cloned Tp.acidophilum genes (Table 4.2). It has also been compared (Table 4.3) with codon usages of a eukaryotic citrate synthase gene, a eubacterial citrate synthase gene and that of three archaebacterial genes.

4.3. Discussion

In the absence of direct proof, no specific conclusions can be drawn on the regulatory regions of the Tp.acidophilum citrate synthase gene. Nevertheless, on the basis of comparison with the consensus sequences observed in other archaebacterial genes, a putative promoter region, putative transcription termination sequences and a putative ribosome-binding site for the Tp.acidophilum gene have been proposed (Fig. 4.1). In addition, a sequence which may be involved in regulation of gene expression has been identified (Fig. 4.1).

Codon usages often vary between genes of the same organism. For example, strongly expressed genes of

Table 4.1: Codon usage of the citrate synthase gene of Tp.acidophilum

F	TTT	3.	S TCT	3.	Y TAT	6.	C TGT	0.
F	TTC	12.	S TCC	5.	Y TAC	14.	C TGC	0.
L	TTA	0.	S TCA	5.	* TAA	0.	* TGA	1.
L	TTG	1.	S TCG	2.	* TAG	0.	W TGG	4.
L	CTT	5.	P CCT	2.	H CAT	2.	R CGT	0.
L	CTC	8.	P CCC	2.	H CAC	4.	R CGC	0.
L	CTA	1.	P CCA	7.	Q CAA	0.	R CGA	1.
L	CTG	9.	P CCG	6.	Q CAG	10.	R CGG	0.
I	ATT	4.	T ACT	3.	N AAT	3.	S AGT	0.
I	ATC	8.	T ACC	6.	N AAC	12.	S AGC	4.
I	ATA	19.	T ACA	5.	K AAA	4.	R AGA	7.
M	ATG	12.	T ACG	7.	K AAG	26.	R AGG	11.
V	GTT	4.	A GCT	7.	D GAT	10.	G GGT	5.
V	GTC	6.	A GCC	10.	D GAC	9.	G GGC	10.
V	GTA	5.	A GCA	19.	E GAA	12.	G GGA	6.
V	GTG	7.	A GCG	8.	E GAG	18.	G GGG	5.

Table 4.2: Comparison of codon usages for cloned Tp.acidophilum genes

Each codon is preceded by the single-letter abbreviation for its amino acid and followed by the percentage of use within its respective amino acid family. Tp: from Tp.acidophilum; CS: citrate synthase; GDH: glucose dehydrogenase [Bright et al. (1990)]; ORF1: open reading frame upstream of 16S rRNA gene [Ree et al. (1989)]; ORF2: open reading frame determined in this project (Chapter 3). For TpORF2 only the amino acids that are represented in statistically relevant numbers have been included.

	TpCS	TpGDH	TpORF 1	TpORF 2		TpCS	TpGDH	TpORF 1	TpORF 2
F TTT	20	20	38	-	Y TAT	30	21	16	-
F TTC	80	80	62	-	Y TAC	70	79	84	-
L TTA	0	0	0	36	H CAT	33	38	0	-
L TTG	4	4	0	9	H CAC	67	62	100	-
L CTT	21	35	36	18					
L CTC	33	35	29	0	M ATG	100	100	100	-
L CTA	4	0	7	9					
L CTG	38	26	28	27	Q CAA	0	0	0	-
					Q CAG	100	100	100	-
I ATT	13	4	0	14					
I ATC	26	19	20	14	N AAC	80	76	100	-
I ATA	61	77	80	72	N AAT	20	24	0	-
V GTT	18	23	41	-	K AAA	13	33	12	-
V GTC	27	27	24	-	K AAG	87	67	88	-
V GTA	23	17	18	-					
V GTG	32	33	18	-	D GAT	53	61	23	-
					D GAC	47	39	77	-
S TCT	16	6	0	0					
S TCC	26	29	29	40	E GAA	40	32	15	-
S TCA	26	18	14	30	E GAG	60	68	85	-
S TCG	11	12	43	20					
S AGT	0	6	0	10	C TGT	0	25	0	-
S AGC	21	29	14	0	C TGC	0	75	0	-
P CCT	12	23	15	-	R CGT	0	17	0	-
P CCC	12	18	23	-	R CGC	0	9	18	-
P CCA	41	35	15	-	R CGA	5	4	0	-
P CCG	35	24	46	-	R CGG	0	4	0	-
					R AGA	37	35	0	-
T ACT	14	10	14	-	R AGG	58	30	82	-
T ACC	29	5	0	-					
T ACA	24	40	14	-	W TGG	100	100	100	-
T ACG	33	45	71	-					
A GCT	16	21	21	-	G GGT	19	22	13	-
A GCC	23	42	50	-	G GGC	38	39	40	-
A GCA	43	16	14	-	G GGA	23	23	20	-
A GCG	18	21	14	-	G GGG	19	16	27	-

Table 4.3: Comparison of codon usages for:
(1) Tp.acidophilum citrate synthase; (2) E.coli citrate synthase [Ner et al. (1983)]; (3) yeast mitochondrial citrate synthase [Rosenkrantz et al. (1986)];
(4) Sulpholobus solfataricus ß-galactosidase [Rossi (1988)]; (5) Methanobacterium thermoautotrophicum mcr (A, B and C) genes [Brown et al. (1989)]; and
(6) Methanobacterium bryantii glyceraldehyde-3-phosphate dehydrogenase [Fabry et al. (1989)]. Each codon is preceded by the single-letter abbreviation for its amino acid and followed by the percentage of use within its respective amino acid family.

	(1)	(2)	(3)	(4)	(5)	(6)		(1)	(2)	(3)	(4)	(5)	(6)
F TTT	20	39	61	57	20	71	Y TAT	30	38	65	47	15	70
F TTC	80	61	39	43	80	29	Y TAC	70	62	35	53	85	30
L TTA	0	0	37	49	4	31	H CAT	33	33	77	50	4	0
L TTG	4	3	35	5	2	4	H CAC	67	67	23	50	96	100
L CTT	21	0	12	11	22	35							
L CTC	33	11	2	3	40	15	M ATG	100	100	100	100	100	100
L CTA	4	0	11	30	4	4							
L CTG	38	86	4	3	27	12	Q CAA	0	12	100	78	4	11
							Q CAG	100	88	0	22	96	89
I ATT	13	54	68	27	11	23							
I ATC	26	46	24	14	33	18	N AAT	20	35	62	63	4	53
I ATA	61	0	8	59	56	59	N AAC	80	65	38	37	96	47
V GTT	18	42	63	55	25	39	K AAA	13	78	59	48	51	87
V GTC	27	16	4	3	35	13	K AAG	87	22	41	52	49	13
V GTA	23	5	11	29	11	39							
V GTG	32	37	22	13	29	10	D GAT	53	56	74	78	20	70
							D GAC	47	44	26	22	80	30
S TCT	16	28	22	19	8	25							
S TCC	26	28	22	10	12	10	E GAA	40	85	77	63	70	85
S TCA	26	12	22	29	57	45	E GAG	60	15	23	37	30	15
S TCG	11	16	8	0	0	0							
S AGT	0	4	11	23	10	10	C TGT	0	57	100	100	78	75
S AGC	21	12	14	19	14	10	C TGC	0	43	0	0	22	25
P CCT	12	10	26	15	24	47	R CGT	0	67	24	3	14	8
P CCC	12	0	9	12	14	20	R CGC	0	33	5	0	2	8
P CCA	41	16	65	69	61	33	R CGA	5	0	0	0	5	8
P CCG	35	74	0	4	2	0	R CGG	0	0	0	0	0	0
							R AGA	37	0	48	59	22	46
T ACT	14	18	54	47	6	42	R AGG	58	0	24	38	58	31
T ACC	29	57	15	11	36	21							
T ACA	24	11	23	32	57	32	W TGG	100	100	100	100	100	100
T ACG	33	14	8	11	0	5							
A GCT	16	24	49	41	17	43	G GGT	19	61	80	33	53	24
A GCC	23	22	29	11	24	10	G GGC	38	29	14	8	10	10
A GCA	43	19	15	30	57	48	G GGA	23	7	0	50	30	59
A GCG	18	35	7	19	2	0	G GGG	19	4	6	10	6	7

E.coli use codons corresponding to iso-accepting tRNAs present in high concentrations, whereas weakly expressed E.coli genes utilise these codons less frequently [Winnacker (1987)]. The codon usages of the Tp.acidophilum genes (given in Table 4.2) are very similar; this may indicate that these three genes are expressed at analogous levels.

The codon usages of the Tp.acidophilum, eubacterial, eukaryotic and other archaebacterial genes shown in Table 4.3 appear to be very different. Nevertheless, it appears that the codon usage of the Tp.acidophilum gene resembles that of the other archaebacterial genes more closely than it does the codon usages of the eubacterial and eukaryotic genes (Table 4.3). No definite conclusions can be drawn from this comparison as the sample used was small, and the method of analysis was purely subjective.

It is perhaps interesting to note that five codons [ATA (ile); AGA (arg); CGA (arg); GGA (gly); and GGG (gly)] infrequently used in E.coli genes [Winnacker (1987) and the E.coli citrate synthase codon usage shown in Table 4.3] are more frequently used in the Tp.acidophilum genes. This does not seem to have affected the efficient translation of the Tp.acidophilum citrate synthase gene in E.coli (see Chapter 6).

CHAPTER 5: ANALYSIS OF THE Tp.acidophilum CITRATE SYNTHASE AMINO ACID SEQUENCE

5.1. Introduction

The following chapter describes the type of information that has been generated from analysis of the Tp.acidophilum citrate synthase amino acid sequence.

5.1.1. Proteins as phylogenetic markers

An increasing number of archaebacterial protein sequences, becoming available as a result of the cloning of their genes, are now being used for phylogenetic comparison with the equivalent eubacterial and eukaryotic proteins. Although these comparisons have confirmed that the archaebacteria are a phylogenetically distinct group of organisms, some contradictory evidence regarding the relationships between the three urkingdoms has been obtained. Whereas 16S rRNA and some protein sequences, for example elongation factor-Tu (EF-Tu) [Lechner & Böck (1987)], indicate that the archaebacteria are more closely related to the eukaryotes than to the eubacteria (see Fig. 1.4.b), other archaebacterial proteins, for example glutamine synthetase (GS) [Sanangelantoni et al. (1990)], show greater sequence similarity to their eubacterial counterparts. Moreover, some eubacterial and eukaryotic proteins, for example glyceraldehyde-3-phosphate dehydrogenases (GAPDH) and malate dehydrogenases (MDH) [Hensel et al. (1989); Honka et al. (1990)], show remarkable similarity to each other, but are very divergent from their archaebacterial equivalents. Zillig et al. (1989) have proposed that the chimeric nature of the eukaryotic cell may explain the observed similarity between some eukaryotic molecules and their archaebacterial or their eubacterial counterparts. In addition, Hensel et al. (1989) have suggested that horizontal gene transfer between urkingdoms may explain some of the observed disparity in the relationships inferred. It is hoped, that by looking at additional

proteins (e.g. citrate synthase), a clearer understanding of the true relationships between eubacteria, eukaryotes and archaebacteria may be achieved.

5.1.2. Secondary-structure prediction

As the structures of homologous proteins are more strongly conserved than their sequences, it is perhaps preferable to use structural similarity rather than sequence homology as a basis for comparison. In the absence of X-ray data for the structures of proteins, various methods have been developed to predict their secondary structures [i.e. α-helices, ß-strands and turns] from their amino acid sequences. Although these methods have been estimated to have success rates of less than 65%, they can be useful in providing an indication of potential secondary structural features.

The program PREDICT (used in this study) gives a joint prediction based on eight different methods; it assigns helix or sheet if three or more of the methods indicate these, and a turn if two or more indicate a turn [Taylor (1990)].

5.1.3. Thermostable proteins

Proteins isolated from thermophilic organisms generally show greater thermostability than their mesophilic counterparts. Studies attempting to elucidate the structural factors that stabilise these thermophilic proteins have ranged from the analysis of 3D-structures to the comparison of amino acid compositions.

The analysis of 3D-structures of mesophilic and thermophilic proteins has given the best clues as to factors important for thermostability of thermophilic proteins. As a result of such a study, Perutz & Raidt (1975) concluded that thermal stability in ferredoxins and haemoglobins is enhanced by the presence of just a few extra salt bridges and/or hydrogen bonds. Walker et al. (1980), on the other hand, found that the differences between GAPDH of lobster and that of the thermophile,

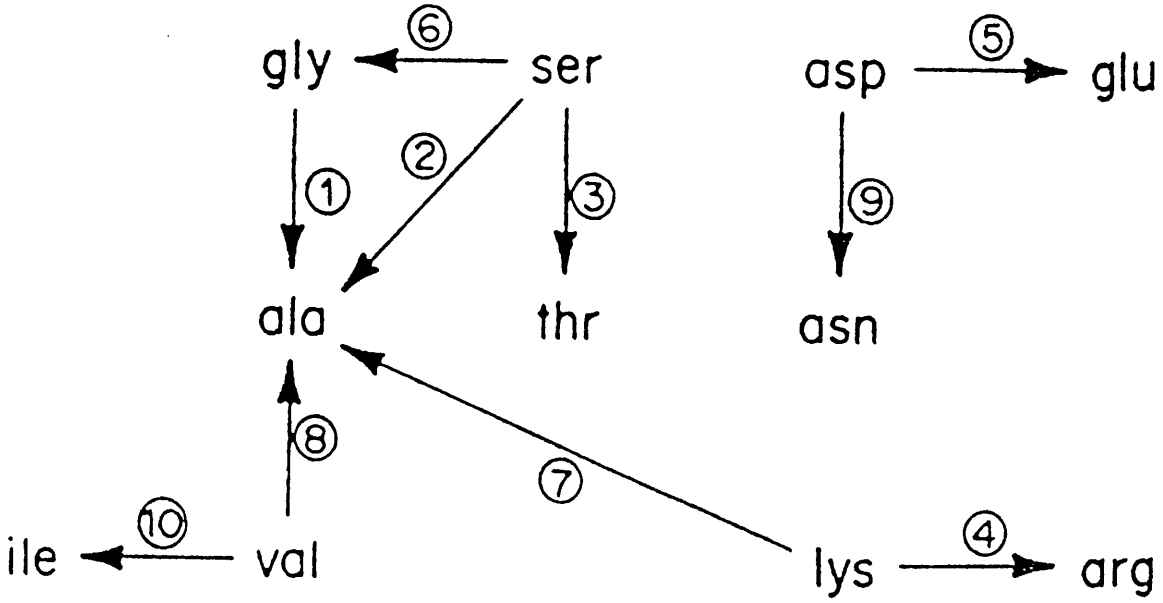
Bacillus stearothermophilus, are quite extensive. Since stability of a protein can be increased by additional stabilisation energy of only 5-10 kcal [Perutz (1978)], the presence of a few extra salt bridges (which contribute 1-3 kcal) can adequately explain the observation of Perutz & Raidt (1975). The requirement, however, that a thermostable protein remains capable of functioning at elevated temperatures, can perhaps explain the finding of Walker et al. (1980); that is, in the case of GAPDH, only a series of small changes has permitted the fine balance necessary between flexibility and desired stability.

Mutagenesis studies have given an insight into some of the amino acid changes that can lead to the enhancement of thermostability of a protein. For example, Grütter et al. (1979) demonstrated that a single exchange (arg → his) in T4 lysozyme was capable of effecting a 14°C decrease in melting temperature, and Ahern et al. (1987) found that replacement of asparagines (asn → thr and asn → ile) in triosephosphate isomerase nearly doubled the half-life of the enzyme at 100°C. Such studies, in addition to establishing rules of thermostability, illustrate possibilities for the redesigning of enzymes for use as industrial catalysts.

Several structural and sequence comparisons between mesophilic and thermophilic proteins [Argos et al. (1979); Alber & Mathews (1987); Crabb et al. (1981); Pauptit et al. (1988); Bowen et al. (1988); Fabry et al. (1989)] have resulted in the observation that various "cold-to-hot" amino acid exchanges are preferred. Fig. 5.1 shows the preferred amino acid exchanges observed by Argos et al. (1979) in their study of GAPDHs, lactose dehydrogenases and ferredoxins. Attempts have been made to explain the observed changes in terms of their effects on the structure and thus the thermostability of the protein. For example, an increase in the proportion of bulkier aliphatic residues

Figure 5.1: Direction of preferred exchanges observed for mesophilic and thermophilic GAPDHs, LDHs and ferredoxins

Reproduced from Argos et al. (1979).
Arrows point from the mesophilic to the thermophilic protein. Numbers indicate the ranking of significance for the given exchange.


gly
ser
asp
glu
ala
thr
asn
ile
val
lys
arg
1
2
3
4
5
6
7
8
9
10

(i.e. gly → ala; and val → ile) has been postulated to enhance stability by increasing the compactness of the folded molecule [Ikai (1980)], and an increase of arginine relative to lysine has been proposed to increase potential for hydrogen bonding [Pauptit et al. (1988)]. Observation of various preferred changes allows a more rational approach to the determination of rules of thermostability by site-directed mutagenesis.

5.2.Results

5.2.1. Alignment of citrate synthase amino acid sequences

The sequence of the Tp.acidophilum citrate synthase was compared with those from several eukaryotes and eubacteria by the computer-aided sequence-alignment method of Henneke et al. (1989) (described in Chapter 2). As the Tp.acidophilum citrate synthase was approximately fifty residues shorter than the other citrate synthases, and the missing amino acids were believed to be the result of a major deletion at the N-terminal of the protein [Smith et al. (1987)], the start of the Tp.acidophilum sequence was aligned approximately fifty residues into the other sequences. The resulting alignment is shown in Fig. 5.2. The sequence identities between each enzyme pair were calculated from this alignment and are presented in Table 5.1.

Table 5.2 lists residues of the Tp.acidophilum citrate synthase that have been identified from the alignment as corresponding to the functionally important residues of the pig enzyme [Remington et al. (1982); Wiegand et al. (1984); Wiegand & Remington (1986); Karpusas et al. (1990)].

5.2.2. Secondary structure of the Tp.acidophilum citrate synthase

The sequences of the Tp.acidophilum and pig citrate synthases were subjected to secondary-structure prediction using the program PREDICT. Fig. 5.3 shows the

Figure 5.2: Alignment of citrate synthase sequences

Sources of the sequence data and order of alignment (top to bottom): (1) pig heart [Bloxham et al. (1981)]; (2) yeast mitochondria [Suissa et al. (1984)]; (3) Arabidopsis thaliana [Unger et al. (1989)]; (4) E.coli [Ner et al. (1983)] - this sequence has been corrected by changing F-228 to V for the reason given in Anderson & Duckworth (1988); (5) Acinetobacter anitratum [Donald & Duckworth (1987)], (6) Pseudomonas aeruginosa [Donald et al. (1989)]; (7) Rickettsia prowazekii [Wood et al. (1987)]; and (8) Tp.acidophilum. Where an amino acid has been conserved between all eight sequences it is highlighted by being repeated on the line directly beneath. Residues of the pig heart enzyme implicated in binding of substrate are marked with a ●, and those believed to be involved in catalysis are marked by a ★. Regions of secondary structure in the pig heart enzyme are indicated by a horizontal line above the sequence.

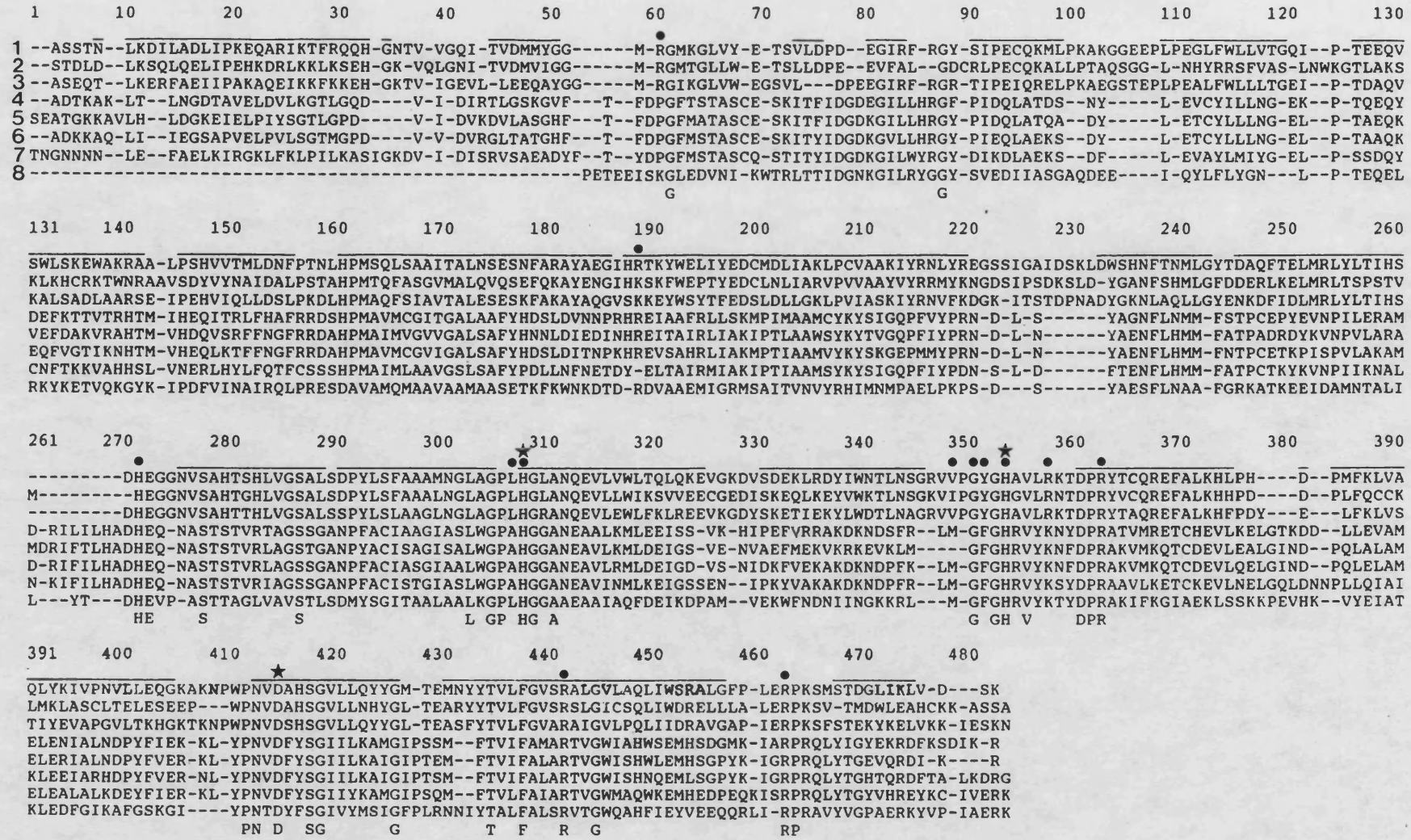

1 10 20 30 40 50 60 70 80 90 100 110 120 130
1 --ASSTN--LKDILADLIPKEQARIKTFRQQH-GNTV-VGQI-TVDMMYGG------M-RGMKGLVY-E-TSVLDPD--EGIRF-RGY-SIPECQKMLPKAKGGEEPLPEGLFWLLVTGQI--P-TEEQV
2 --STDLD--LKSQLQELIPEHKDRLKKLKSEH-GK-VQLGNI-TVDMVIGG------M-RGMTGLLW-E-TSLLDPE--EVFAL--GDCRLPECQKALLPTAQSGG-L-NHYRRSFVAS-LNWKGTLAKS
3 --ASEQT--LKERFAEIIPAKAQEIKKFKKEH-GKTV-IGEVL-LEEQAYGG-----M-RGIKGLVW-EGSVL---DPEEGIRF-RGR-TIPEIQRELPKAEGSTEPLPEALFWLLLTGEI--P-TDAQV
4 --ADTKAK-LT--LNGDTAVELDVLKGTLGQD----V-I-DIRTLGSKGVF---T--FDPGFTSTASCE-SKITFIDGDEGILLHRGF-PIDQLATDS--NY-----L-EVCYILLNG-EK--P-TQEQY
5 SEATGKKAVLH--LDGKEIELPIYSGTLGPD-----V-I-DVKDVLASGHF---T--FDPGFMATASCE-SKITFIDGDKGILLHRGY-PIDQLATQA--DY-----L-ETCYLLLNG-EL--P-TAEQK
6 --ADKKAQ-LI--IEGSAPVELPVLSGTMGPD----V-V-DVRGLTATGHF---T--FDPGFMSTASCE-SKITYIDGDKGVLLHRGY-PIEQLAEKS--DY-----L-ETCYLLLNG-EL--P-TAAQK
7 TNGNNNN--LE--FAELKIRGKLFKLPILKASIGKDV-I-DISRVSAEADYF--T--YDPGFMSTASCQ-STITYIDGDKGILWYRGY-DIKDLAEKS--DF-----L-EVAYLMIYG-EL--P-SSDQY
8 ----------PETEEISKGLEDVNI-KWTRLTTIDGNKGILRYGGY-SVEDIIASGAQDEE----I-QYLFLYGN---L--P-TEQEL
G G
131 140 150 160 170 180 190 200 210 220 230 240 250 260
SWLSKEWAKRAA-LPSHVVTMLDNFPTNLHPMSQLSAAITALNSESNFARAYAEGIHRTKYWELIYEDCMDLIAKLPCVAAKIYRNLYREGSSIGAIDSKLDWSHNFTNMLGYTDAQFTELMRLYLTIHS
KLKHCRKTWNRAAVSDYVYNAIDALPSTAHPMTQFASGVMALQVQSEFQKAYENGIHKSKFWEPTYEDCLNLIARVPVVAAYVYRRMYKNGDSIPSDKSLD-YGANFSHMLGFDDERLKELMRLTSPSTV
KALSADLAARSE-IPEHVIQLLDSLPKDLHPMAQFSIAVTALESESKFAKAYAQGVSKKEYWSYTFEDSLDLLGKLPVIASKIYRNVFKDGK-ITSTDPNADYGKNLAQLLGYENKDFIDLMRLYLTIHS
DEFKTTVTRHTM-IHEQITRLFHAFRRDSHPMAVMCGITGALAAFYHDSLDVNNPRHREIAAFRLLLSKMPIMAAMCYKYSIGQPFVYPRN-D-L-S------YAGNFLNMM-FSTPCEPYEVNPILERAM
VEFDAKVRAHTM-VHDQVSRFFNGFRRDAHPMAIMVGVVGALSAFYHNNLDIEDINHREITAIRLIAKIPTLAAWSYKYTVGQPFIYPRN-D-L-N------YAENFLHMM-FATPADRDYKVNPVLARA
EQFVGTIKNHTM-VHEQLKTFFNGFRRDAHPMAVMCGVIGALSAFYHDSLDITNPKHREVSAHRLIAKMPTIAAMVYKYSKGEPMMYPRN-D-L-N------YAENFLHMM-FNTPCETKPISPVLAKAM
CNFTKKVAHHSL-VNERLHYLFQTFCSSSHPMAIMLAAVGSLSAFYPDLLNFNETDY-ELTAIRMIAKIPTIAAMSYKYSIGQPFIYPDN-S-L-D------FTENFLHMM-FATPCTKYKVNPIIKNAL
RKYKETVQKGYK-IPDFVINAIRQLPRESDAVAMQMAAVAAMAASETKFKWNKDTD-RDVAAEMIGRMSAITVNVYRHIMNMPAELPKPS-D---S------YAESFLNAA-FGRKATKEEIDAMNTALI
261 270 280 290 300 310 320 330 340 350 360 370 380 390
---------DHEGGNVSAHTSHLVGSALSDPYLSFAAAMNGLAGPLHGLANQEVLVWLTQLQKEVGKDVSDEKLRDYIWNTLNSGRVVPGYGHAVLRKTDPRYTCQREFALKHLPH----D--PMFKLVA
M---------HEGGNVSAHTGHLVGSALSDPYLSFAAALNGLAGPLHGLANQEVLLWIKSVVEECGEDISKEQLKEYVWKTLNSGKVIPGYGHGVLRNTDPRYVCQREFALKHHPD----D--PLFQCCK
---------DHEGGNVSAHTTHLVGSALSSPYLSLAAGLNGLAGPLHGRANQEVLEWLFKLREEVKGDYSKETIEKYLWDTLNAGRVVPGYGHAVLRKTDPRYTAQREFALKHFPDY---E---LFKLVS
D-RILILHADHEQ-NASTSTVRTAGSSGANPFACIAAGIASLWGPAHGGANEAALKMLEEISS-VK-HIPEFVRRAKDKNDSFR--LM-GFGHRVYKNYDPRATVMRETCHEVLKELGTKDD--LLEVAM
MDRIFTLHADHEQ-NASTSTVRLAGSTGANPYACISAGISALWGPAHGGANEAVLKMLDEIGS-VE-NVAEFMEKVKRKEVKLM-----GFGHRVYKNFDPRAKVMKQTCDEVLEALGIND--PQLALAM
D-RIFILHADHEQ-NASTSTVRLAGSSGANPFACIASGIAALWGPAHGGANEAVLRMLDEIGD-VS-NIDKFVEKAKDKNDPFK--LM-GFGHRVYKNFDPRAKVMKQTCDEVLQELGIND--PQLELAM
N-KIFILHADHEQ-NASTSTVRIAGSSGANPFACISTGIASLWGPAHGGANEAVINMLKEIGSSEN--IPKYVAKAKDKNDPFR--LM-GFGHRVYKSYDPRAAVLKETCKEVLNELGQLDNNPLLQIAI
L---YT---DHEVP-ASTTAGLVAVSTLSDMYSGITAALAALKGPLHGGAAEAAIAQFDEIKDPAM--VEKWFNDNIINGKKRL---M-GFGHRVYKTYDPRAKIFKGIAEKLSSKKPEVHK--VYEIAT
HE S S L GP HG A G GH V DPR
391 400 410 420 430 440 450 460 470 480
QLYKIVPNVLLEQGKAKNPWPNVDAHSGVLLQYYGM-TEMNYYTVLFGVSRALGVLAQLIWSRALGFP-LERPKSMSTDGLIKLV-D---SK
LMKLASCLTELESEEP---WPNVDAHSGVLLNHYGL-TEARYYTVLFGVSRSLGICSQLIWDRELLLA-LERPKSV-TMDWLEAHCKK-ASSA
TIYEVAPGVLTKHGKTKNPWPNVDSHSGVLLQYYGL-TEASFYTVLFGVARAIGVLPQLIIDRAVGAP-IERPKSFSTEKYKELVKK-IESKN
ELENIALNDPYFIEK-KL-YPNVDFYSGIILKAMGIPSSM--FTVIFAMARTVGWIAHWSEMHSDGMK-IARPRQLYIGYEKRDFKSDIK-R
ELERIALNDPYFVER-KL-YPNVDFYSGIILKAIGIPTEM--FTVIFALARTVGWISHWLEMHSGPYK-IGRPRQLYTGEVQRDI-K----R
KLEEIARHDPYFVER-NL-YPNVDFYSGIILKAIGIPTSM--FTVIFALARTVGWISHNQEMLSGPYK-IGRPRQLYTGHTQRDFTA-LKDRG
ELEALALKDEYFIER-KL-YPNVDFYSGIIYKAMGIPSQM--FTVLFAIARTVGWMAQWKEMHEDPEQKISRPRQLYTGYVHREYKC-IVERK
KLEDFGIKAFGSKGI----YPNTDYFSGIVYMSIGFPLRNNIYTALFALSRVTGWQAHFIEYVEEQQRLI-RPRAVYVGPAERKYVP-IAERK
PN D SG G T F R G RP

Table 5.1: Sequence identities between eukaryotic, eubacteria and archaebacterial citrate synthases

citrate synthase source	% identity with citrate synthase from: (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) pig	100	60	50	23	23	21	19	20
(2) Yeast		100	48	22	22	20	19	20
(3) *A. thaliana*			100	19	21	19	18	19
(4) *E. coli*				100	62	69	57	27
(5) *A. anitratum*					100	70	56	28
(6) *Ps. aeruginosa*						100	57	27
(7) *R. prowazekii*							100	28
(8) *Tp. acidophilum*								100

Table 5.2: Predicted active site residues in pig and Tp.acidophilum citrate synthases

Function	Residue in pig heart citrate synthase	Equivalent residue in Tp.acidophilum citrate synthase	Reference(s) (see *)
Binding of oxaloacetate and/citrate (H-bonds)	His-238	His-187	(1),(2),(3)
	His-274	His-222	"
	His-320	His-262	"
	Arg-329	Arg-271	"
	Arg-401	Arg-344	"
	Arg-421	Arg-364	"
Binding of Co-A (H-bonds)	Leu-273	Leu-221	(1)
	Val-315	Met-258	(1),(3)
	Gly-317	Gly-259	"
	Tyr-318	Phe-260	"
	Arg-164	Arg-121	"
	Arg-46	Lys-8	"
	Lys-368 (trimethyl)	-	(1)
Binding CoA 3'-phosphate	Arg-164	Arg-121	(1),(3)
Binding CoA 5'-diphosphate	Arg-46	Lys-8	(1),(3)
	Arg-324	Lys-266	"
Catalytic	His-274	His-222	(1),(3),(4)
	His-320	His-262	"
	Asp-375	Asp-317	"

* References: (1) Remington et al. (1982); (2) Wiegand et al. (1984); (3) Wiegand & Remington (1986); (4) Karpusas et al. (1990).

Figure 5.3: Secondary-structure predictions for pig and Tp.acidophilum citrate synthases

The alignment of pig heart citrate synthase with that of Tp.acidophilum was taken from the multiple alignment shown in Fig. 5.2.

1: Tp.acidophilum citrate synthase sequence
2: Pig heart citrate synthase sequence
3: Predicted secondary structure of Tp.acidophilum CS
4: Predicted secondary structure of pig CS
5: Actual secondary structure of pig CS

The α-helices of the pig enzyme are labelled (a) to (t).

H or h: α-helices; B or b: ß-strands; and T or t: turns [upper-case has been used for actual secondary structure and lower-case for predicted secondary structures].

```
                (a)                          (b)
1                                                    PETEEISKGLEDVNIKWTRL
2    ASSTNLKDILADLIPKEQARIKTFRQQHGNTVVGQITVDMMYGG-----M-RGMKGLVYE-TSV
3                                                   hhhhhhhhhhhhhhbbbbbb
4    ttt h hhhhhhhhhhhhhhh     tttbbbbbbbbbbtttt     t tttbbbbbb bbb
5         THHHHHHHHHHHHHHHHHHHHHTTT       HHHHHTT     T TT        T B

                      (c)                (d)              (e)
1    TTIDGNKGILRYGGYSVEDIIASGAQDEE----I-QYLFLYGN---LPTEQELRKYKETVQKGY
2    LDPD--EGIRF-RGYSIPECQKMLPKAKGGEEPLPEGLFWLLVTGQIPTEEQVSWLSKEWAKRA
3    bb ttttbbbbtttttbbbbbb thhhhh     h hhh bb t   tthhhhhhhhhhhh ttt
4    bttt  tttbb ttttt  hhhhh ttttttttttthhhhbbbbbbb thhhh hhhhhhhhhh
5    BBTT  TBBBT TT BHHHHHHHT B TT TTB HHHHHHHHHTTT   HHHHHHHHHHHHHH

          (f)             (g)             (h)              (i)
1    KIPDFVINAIRQLPRESDAVAMQMAAVAAMAASETKFKWNKDTD-RDVAAEMIGRMSAITVNVY
2    ALPSHVVTMLDNFPTNLHPMSQLSAAITALNSESNFARAYAEGIHRTKYWELIYEDCMDLIAKL
3    ttbbbbbb hhhttthhhhhhhhhhhhhhhhhhhhhhhttttt hhhhhhhhhhhh bbbbbb
4    hht bbbbb  tttttt  thhhhhhhhhhh tthhhhhhhhhh     hhhhhhhhhhhhhhhh
5       HHHHHHHHTT TTT HHHHHHHHHHHTTTTT HHHHHHHHT  TTTHHHHHHHHHHHHHHHH

                                    (j)          (k)
1    RHIMNMPAELPKPS-D---S------YAESFLNAA-FGRKATKEEIDAMNTALILYTDHEVP-A
2    PCVAAKIYRNLYREGSSIGAIDSKLDWSHNFTNMLGYTDAQFTELMRLYLTIHS---DHEGGNV
3    bbbb  at ttttt t   t       hhhhhhhh hhhhhhhhhhhhhhhhhhbbbbtt tt t
4    h hhhhhhhbb tttttt  b t  tt thhbb   thhhhhhhhhhh  bbh    httttbb
5    HHHHHHHHHHHHTT T       TTT HHHHHHHHHT   HHHHHHHHHHHHHTT      T TH

       (l)             (m)                  (n)                  (o)
1    STTAGLVAVSTLSDMYSGITAALAALKGPLHGGAAEAAIAQFDEIKDPAM--VEKWFNDNIING
2    SAHTSHLVGSALSDPYLSFAAAMNGLAGPLHGLANQEVLVWLTQLQKEVGKDVSDEKLRDYIWN
3    tt bbbbbbbbbbttttthhhhhhhhhhttttttthhhhhhhhhhhhhhhhhh  hhhhhht b hh
4    b btbbbb ttttttt hhhhhhhttttthhhhhhhhhhhhhhhhhhhhhhhhhhhhh bbbb
5    HHHHHHHTTTTT  HHHHHHHHHHHHHTTTTTT THHHHHHHHHHHHHHHT TT  TTHHHHHHHH

                          (p)                       (q)
1    KKRL---M-GFGHRVYKTYDPRAKIFKGIAEKLSSKKPEVHK-VYEIATKLEDFGIKAFGSKGI
2    TLNSGRVVPGYGHAVLRKTDPRYTCQREFALKHLPH----D-PMFKLVAQLYKIVPNVLLEQGK
3    hhh   h  tttbbbbbtthhhhhhhhhhhhhhht hhhhhh hhhhhhhhhhhhhh  ttttt
4    btttt bbbttbbbbbbttttt thhhhhhhhhtt       hhhhhhhhh bbbbbbhhhhht
5    HHHTT   TTB  TT  T  HHHHHHHHHHHHHHTTT      HHHHHHHHHHHHHHHHHHHHTT

           (r)                  (s)
1    ----YPNTDYFSGIVYMSIGFPLRNNIYTALFALSRVTGWQAHFIEYVEEQQRLI-RPRAVYVG
2    AKNPWPNVDAHSGVLLQYYGM-TEMNYYTVLFGVSRALGVLAQLIWSRALGFP-LERPKSMSTD
3      tttttt bbbbbbbbttttt bbbbbbbbbbbbbt  hhhhhhhhhhhhhb bbbbbbbbtt
4     ttttttt tttbbbbbbbbb bbbbbbbbbbbbbhhhhhhhhhhhhhh  tt ttttttttt
5     T   B THHHHHHHHHTTT    TTTHHHHHHHHHHHHHHHHHHHHHHTT          HH

      (t)
1    PAERKYVPIAERK
2    GLIKLV-D--SK
3    ttt  bbbb hhh
4    hbbbh    tt
5    HHHHHH T
```

aligned Tp.acidophilum and pig citrate synthase sequences, the secondary structure predicted for both enzymes and the secondary structural regions of the pig enzyme derived from X-ray data. Table 5.3 lists the actual α-helices of the pig enzyme and gives an indication of those correctly predicted.

5.2.3 Amino acid exchanges between citrate synthases

The amino acid changes between (1) the Tp.acidophilum and pig citrate synthases, and between (2) the Tp.acidophilum and E.coli citrate synthases are shown in Tables 5.4 and 5.5, respectively. Net amino acid changes for enzyme pairs (1) and (2) were then calculated; these are shown in Table 5.6 and 5.7, respectively. The most striking changes observed in each of the two comparisons have been listed in Table 5.8, along with the corresponding changes observed between pig and E.coli (the latter are included as a control).

5.3. Discussion

5.3.1. Alignment of citrate synthase amino acid sequences

It is quite clear that the eukaryotes, the eubacteria and Tp.acidophilum fall into three separate groups since there is a high degree of identity between citrate synthases within a kingdom (i.e. the eukaryotic and the eubacterial citrate synthases have mean identities of 53% and 62%, respectively), whereas lower identities (i.e. 18-28%) are observed between the kingdoms (Table 5.1). The Tp.acidophilum enzyme is significantly more related to the eubacterial citrate synthase (mean identity 28%) than it is to the eukaryotic enzyme (mean identity 20%) and than the eubacterial enzyme is to the eukaryotic enzyme (mean identity 21%). So, although providing evidence in support of Woese's archaebacterial tree, the data do not disprove Lake's eocyte tree, in which Tp.acidophilum (along with the methanogens and halophiles) is grouped more closely to

Table 5.3: Percent α-helical residues of pig citrate synthase correctly predicted (by PREDICT)

Actual α-helix in pig CS	Residues (%) correctly predicted in pig CS
A (7-27)	76
B (38-42)	0
C (71-77)	57
D (89-97)	44
E (104-117)	93
F (122-129)	0
G (137-146)	80
H (153-160)	100
I (167-193)	89
J (209-217)	22
K (222-234)	69
L (243-250)	0
M (258-269)	58
N (278-291)	100
O (300-310)	27
P (328-340)	62
Q (345-363)	63
R (375-383)	0
S (393-414)	64
T (427-434)	25

(For example, 76% of residues 7-27 of helix A were correctly predicted)

Table 5.4: Substitution of residues between Tp.acidophilum and pig heart citrate synthase sequences

Tp.acidophilum and pig heart sequences were aligned as shown in Fig. 5.2. Underlined values along the diagonal indicate the conserved residues. All other values indicate the type of replacement (e.g. two of the four pig cysteine residues have been changed to two isoleucines, and the other two to a serine and a histidine).

Residue of pig CS	of Tp.acidophilum																					
	G	A	V	L	I	F	M	C	P	Y	T	S	W	Q	N	D	E	H	R	K	Absent	Total
G	11	2	2				1		2	1	1			1		1	1				10	33
A		8	1			1	1			2	4	1	1	1	2	1	1		1	4	4	33
V	1	8	3	3	3		1			1	1		1		2						4	28
L	4	3	3	6	7	3	2		1	5	3	3		2	2		3			2	4	53
I		1	3	1	4	1			1							1	1				6	19
F				1	2	3				1	1			1					1	1	1	12
M		3	3	2	2	1						1			1						2	15
C					2							1						1				4
P		2			2		1		7			1					1		2		6	22
Y	1	3			4	2	1			3		1			1		1			1	1	19
T	1	2	1	2						1	4				1	2			1	3	5	23
S	2	4		1	1		1			2		6				2	3		2	1	4	29
W		1	1	1						2				1	1		1			1		9
Q		1			1	1	2										3	1		3	5	17
N	1	3		1							1	1		1	3		1		1	1	4	18
D	1				1		1		1		1				1	5		1	2	1	6	21
E	1	1							1		1	3	1	2		2	4		2	2	4	24
H				3		2					1					1	1	3		1	2	14
R	2		1	1					1							2	1		4	4	3	19
K		1	2		1	1	1				1			1		2	4		3	2	6	25
Absent	1	1	2	2	1				3	2	2	1	1		1		4			3		
Total	26	44	22	24	31	15	12	0	17	20	21	19	4	10	15	19	30	6	19	30		

Table 5.5: Substitution of residues between Tp.acidophilum and E.coli sequences

Tp.acidophilum and E.coli sequences were aligned as shown in Fig. 5.2. Underlined values along the diagonal indicate the conserved residues. All other values indicate the type of replacement (for an example see Table 5.4)

Residue of E.coli CS	Residue of Tp.acidophilum CS																					
	G	A	V	L	I	F	M	C	P	Y	T	S	W	Q	N	D	E	H	R	K	Absent	Total
G	13	3	1	1			1		1					1			1				6	28
A	1	18	1	2	1						1	4		1	3						5	37
V	1	2	2	2	3		1			1	2				1						4	19
L	2	1	2	4	5	1	1			4		1			1				1	2	10	35
I		1	6	4	9	2	1				1			1	1					1	3	30
F				2	2	5				6	1	2	1				2		1		1	23
M		2	2	1	2	1	1			1	2	1		3	1					1		18
C	1	2		1			1			1											1	7
P		4				1	1		6		1	1					1			4		19
Y	1			1	1	2			2	5							3		1			16
T		2	4		2					1	6			1			3		1	1	5	26
S	2	1		1	2	1	1		1		3	5				2	1		2	2	1	25
W						1							1							1		3
Q		1	1			1			1							1	2				1	8
N	1	1									1	2			3	5				2	2	17
D		1		1	1							2	1		1	5	3	1	1	4	6	27
E	2	1									2			2		3	7			7	2	26
H	1		1						1	1	1					1	1	4	1		3	15
R	1	1		2			1		2						2	2			9	2	2	24
K		2	2	1	2		3					1			1		1	1	2	2	5	23
Absent		1		1	1				3				1	1	1		5			1		
Total	26	44	22	24	31	15	12	0	17	20	21	19	4	10	15	19	30	6	19	30		

Table 5.6: Net amino acid exchanges between Tp.acidophilum and pig citrate synthases

The data have been derived from Table 5.4. For example, 8 valines of the pig CS have changed to alanines in the Tp.acidophilum enzyme, whereas, in the reverse direction, only 1 valine of the Tp.acidophilum CS has changed to an alanine (Table 5.4). This means there is a net valine to alanine change of 7 from pig to Tp.acidophilum (alternatively it can be called a net alanine to valine change of 7 from Tp.acidophilum to pig) and this figure is indicated in the table.

Residue of pig CS	Residue of Tp.acidophilum CS																			
	G	A	V	L	I	F	M	C	P	Y	T	S	W	Q	N	D	E	H	R	K
G		2	1				1		2					1						
A						1					2					1			1	3
V		7								1					2					
L	4	3			6	2			1	5	1	2		2	1		3			2
I		1															1			
F					1						1								1	
M		2	2		2	1									1					
C					2							1						1		
P		2			1		1					1							1	
Y		1			4	1	1								1		1			1
T										1						1			1	2
S	2	3			1					1						2			2	1
W				1						2				1	1					1
Q					1		2										1	1		2
N	1	1										1		1			1		1	1
D							1		1						1					
E											1					2			1	
H				3		2					1						1			1
R	2		1	1																1
K			2		1		1									1	2			

Table 5.7: Net amino acid exchanges between Tp.acidophilum and E.coli citrate synthases

The data have been derived from Table 5.5. For example, 1 alanine of the E.coli CS has changed to glycine in the Tp.acidophilum enzyme, whereas, in the reverse direction, 3 alanines of the Tp.acidophilum CS have changed to glycines (Table 5.5). This means there is a net alanine to glycine change of 2 from Tp.acidophilum to E.coli (alternatively it can be looked at as a net glycine to alanine change of 2 from E.coli to Tp.acidophilum) and this figure is indicated in the table.

Residue of E.coli CS	Residue of Tp.acidophilum CS																			
	G	A	V	L	I	F	M	C	P	Y	T	S	W	Q	N	D	E	H	R	K
G		2					1		1					1						
A				1								3			2					
V		1								1					1					
L	1				1					3					1					1
I			3											1	1					
F				1						4	1	1					2		1	
M		2	1		1	1				1	2			3	1					
C	1	2		1			1			1										
P		4				1	1				1						1			4
Y	1				1				2								3		1	
T		1	2		1					1				1			1		1	1
S	2				2						3						1		2	1
W																				1
Q			1			1			1							1				
N	1										1	2				4				1
D		1		1	1								1							4
E	1	1																		6
H	1		1						1	1	1						1		1	
R	1	1		1			1		2						2	1				
K		2	2		1		2											1		

Table 5.8: Observed preferred exchanges among Tp.acidophilum, E.coli and pig citrate synthases

Data presented here have been selected from Tables 5.6 and 5.7. Positive numbers show that the change is in the direction indicated (i.e. E.coli to Tp.acidophilum for column 1), whereas negative numbers show that the change is in the opposite direction to that indicated (i.e. Tp.acidophilum to E.coli for column 1)

Amino acid exchanges	Net citrate synthase amino acid exchange from		
	E.coli to Thermoplasma	Pig to Thermoplasma	Pig to E.coli
Val to ala	+1	+7	+6
Leu to ile	+1	+6	+3
Glu to lys	+6	-2	-2
Leu to tyr	+3	+5	+2
Pro to lys	+4	0	+2
Asn to asp	+4	-1	0
Asp to lys	+4	-1	+1
Leu to gly	+1	+4	+2
Phe to tyr	+4	-1	-4
Tyr to ile	+1	+4	+1
Pro to ala	+4	+2	0

the eubacteria.

As the closer similarity between the eubacterial and archaebacterial citrate synthases is unlikely to be the result of sequence convergence, this result implies that these citrate synthases resemble that of the universal ancestor more closely than the eukaryotic form of the enzyme does. Whether this relationship between citrate synthases is a true reflection of the relationship between the urkingdoms is unclear since, as pointed out earlier (in Section 5.1.1), the relationship inferred seems to depend upon which protein (e.g. EF-Tu, GS, or MDHase) is chosen as the phylogenetic marker.

The percent identities observed in this study are in basic agreement with those found for citrate synthases by Donald et al. (1989); however the figures quoted by the latter were 1-5% greater than those observed here. This result was not unexpected since the alignment of Donald et al. (1989) was optimised to give maximum identities, whereas the alignment proposed here incorporated penalties for gaps and insertions in regions of secondary structure.

Despite the low sequence identity between the pig and the Tp.acidophilum citrate synthases, 10 of the 15 residues implicated in either binding of the substrates or the catalytic mechanism of the pig enzyme appear to be unchanged in the Tp.acidophilum enzyme. Three of the remaining five catalytically-important residues show conservative changes: tyrosine to phenylalanine or arginine to lysine (Table 5.2, & Fig. 5.2).

5.3.2 Citrate synthase secondary-structure prediction

The secondary structure of the Tp.acidophilum citrate synthase was predicted (Fig. 5.3) using the program PREDICT. In order to estimate the accuracy of this method, a prediction for the pig citrate synthase was also carried out and compared to the actual structure derived from X-ray data (Fig. 5.3).

The majority of the α-helices of the pig citrate synthase were correctly assigned in the prediction for the pig enzyme, although the lengths of these varied considerably from those actually present (Table 5.3). The predictions for both the pig and Tp.acidophilum enzymes suggest extensive regions of ß-sheet, whereas the X-ray data for the pig enzyme indicate that there is only one such region. Overall, only 46% of the pig residues were correctly assigned, suggesting that methods for secondary-structure prediction are not very accurate. The prediction for the Tp.acidophilum citrate synthase should therefore be viewed with caution.

Two α-helices were predicted at the N-terminal of the Tp.acidophilum citrate synthase, whereas no such helices occur in the corresponding region of the pig enzyme (Fig. 5.3). In a slightly different sequence alignment (e.g. where no account is taken that the Tp.acidophilum citrate synthase is 50 amino acids shorter than the pig enzyme), the two additional helices of the Tp.acidophilum enzyme could conceivably correspond to helices (a) and (b) of the pig protein (Fig. 5.3). This would suggest that the major deletion of the Tp.acidophilum enzyme had not occurred directly at the N-terminal as proposed by Smith et al. (1987). It is difficult, however, to envisage how a deletion could have occurred elsewhere in the Tp.acidophilum citrate synthase protein, since alternative regions are likely to be involved in catalysis and/or subunit-subunit interactions (see Fig. 1.2). Moreover, the secondary-structure prediction was viewed as unreliable (see above), and so the deletion proposed by Smith et al. (1987) is still considered the most probable.

5.3.3. Amino acid exchanges between citrate synthases

The preferred amino acid exchanges between the Tp.acidophilum and pig citrate synthases are very different from those observed between the Tp.acidophilum and E.coli citrate synthases (Table 5.8). In addition,

they do not appear to agree with the preferred changes observed for other thermostable proteins (see Fig. 5.1). This finding was not unexpected since (as discussed in Section 5.3.1) pig, E.coli and Tp.acidophilum are very different phylogenetically, and a larger number of amino acid changes would have been as a result of the great evolutionary distance between the organisms. Nevertheless, it is worth noting two features of the Tp.acidophilum citrate synthase sequence which may be connected with the thermostability of the enzyme, namely: (1) a lack of cysteines, a feature observed in other thermophilic enzymes [Hocking & Harris (1976)]; and (2) an increase in alanines, a trait which has been proposed to afford greater alpha-helical strength and hydrophobicity [Argos et al. (1979)].

The phylogenetic and the thermostability comparisons, discussed in this chapter, have emphasised a need for citrate synthase sequences from other archaebacterial species (e.g. a halophile and a sulphur-dependent thermophile). Phylogenetic comparisons, in which these additional sequences are included, may provide evidence to support Woese's (or Lake's) phylogenetic tree. In addition, sequence comparisons with the archaebacterial citrate synthases [assuming that they prove to be as similar to each other as the eukaryotic (mean identity 53%) and eubacterial (mean identity 62%) citrate synthases] may allow a more conclusive analysis of factors enhancing thermostability of the Tp.acidophilum citrate synthase.

The lack of success in predicting secondary structure has indicated that a 3D-structure of the Tp.acidophilum citrate synthase derived from X-ray data is required. Early stages towards the achievement of this aim will be discussed in Chapters 6 and 7.

CHAPTER 6: EXPRESSION OF *Thermoplasma acidophilum* CITRATE SYNTHASE IN *Escherichia coli*

6.1. Introduction

The aim of the experiments described in this chapter was to obtain expression in *E.coli* of a functional *Tp.acidophilum* citrate synthase from pUC constructs of the gene.

For successful gene expression, regulatory regions controlling initiation and termination of both transcription and translation are required. If such regions, compatible with the RNA polymerase and ribosome of the host, are not present on the cloned gene, it may be possible to introduce them on the cloning vector.

The promoter of a gene, a region recognised by the RNA polymerase, is required for initiation of transcription. In *E.coli*, the promoter consists of two consensus sequences at regions "-10"(TATAAT) and "-35"(TTGACA) [Pribnow (1975)]. This is different from the consensus determined for archaebacterial promoters [Zillig *et al.* (1988)] and that proposed for the *Tp.acidophilum* citrate synthase gene (discussed in Chapter 4). In order to ensure successful transcription in this study, the inducible *E.coli lac* promoter on the pUC vector was employed.

Initiation of translation requires a ribosome-binding site. The following rules, discussed by Winnacker (1987), have been proposed for the *E.coli* ribosome-binding site:

Rule I: The typical primary structure is $AGGN_{6-9}ATG$
Rule II: G should not occur at position "-3"
Rule III: Less than two G residues should occur between positions "-1" and "-7"
Rule IV: Positions "+5" and "+10" should be occupied by either A or T

Except for the occurrence of a C in position "+5", the putative ribosome-binding site of the Tp.acidophilum citrate synthase gene (discussed in Chapter 4) adheres to all the above rules (see Fig. 4.1). It was therefore predicted that the E.coli ribosome would initiate translation from the Tp.acidophilum ribosome-binding site. A ß-galactosidase fusion protein, the result of initiation of translation from the lac ribosome-binding site of pUC, was viewed as undesirable.

The codon preferences of different organisms can be very variable and this can be a potential problem for the expression of a foreign gene in an heterologous organism. As the codon usage observed in the Tp.acidophilum citrate synthase gene is different from that found in E.coli genes (discussed in Chapter 4), this was considered a potential problem for its efficient expression in E.coli.

6.2. Results

6.2.1. Construction of plasmid pTaCS18

In pTaCS19 the citrate synthase gene was in the correct orientation required for expression from the lac promoter and lac ribosome-binding site (Fig. 6.1). A second construct, designated pTaCS18, was derived from pTaCS1 and pTaCS2 in the manner described for the construction of pTaCS19 (Section 3.2.8), but pUC18 was used in place of pUC19 at step (a) in Fig. 3.16. Digestion of pTaCS18 with PvuII (Fig. 6.2, Table 6.1) and its partial sequencing (Fig. 6.3.d) confirmed that the citrate synthase gene was present, and in the opposite orientation to that of pTaCS19.

6.2.2. Construction of plasmids pCSEH18 and pCSEH19

In pTaCS18 and pTaCS19 the citrate synthase gene was preceded by approximately 1-kb of "non-coding" Tp.acidophilum DNA. This 1-kb fragment was considered potentially disruptive to optimal transcription of the gene, and was removed by the scheme described below and

Figure 6.1: Construction of plasmids pCSEH18, pCSEH19 and pCSSS

(a) pTaCS19, digested with HincII and EcoRV, was religated to generate the clone pCSEH19. (b) pTaCS18, digested with HincII and EcoRV, was religated to generate the clone pCSEH18. (c) pCSEH19 was digested with SphI, treated with S1 nuclease, and religated to generate the clone pCSSS.

The lac control region and the 5′ end of the lacZ gene (carried on the pUC vector) are unshaded. An arrow indicates the direction of transcription for lacZ. Tp.acidophilum DNA, prior to the start of the citrate synthase gene, is represented by the hatched region. The citrate synthase gene is represented by the shaded region.

e = EcoRI; E = EcoRV; H = HincII; P = PvuII; S = SphI.

(N.B. not to scale).

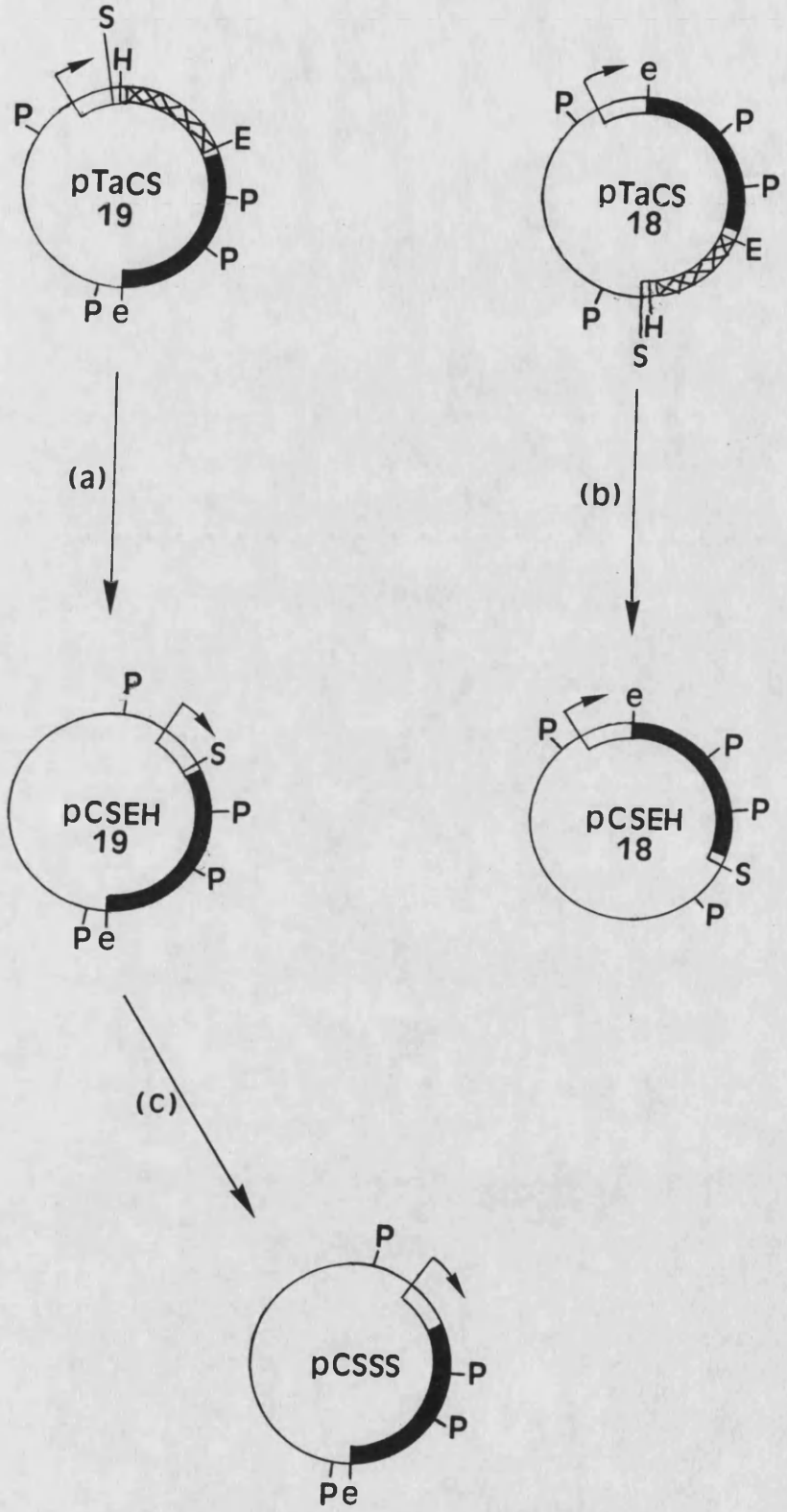

S
H
P
E
pTaCS
19
P
P
P e
e
P
P
P
E
pTaCS
18
P
H
S
(a)
(b)
P
S
pCSEH
19
P
P
P e
e
P
P
pCSEH
18
S
P
(c)
P
pCSSS
P
P
P e

Figure 6.2: PvuII-analysis of pTaCS18, pTaCS19, pCSEH18 and pCSEH19

An agarose gel (1.2%) showing PvuII digests of the plasmids pTaCS18 (track 1), pCSEH18 (track 2), pTaCS19 (track 3) and pCSEH19 (track 4). Sizes of Lambda DNA HindIII-digested fragments are indicated in the left-hand margin. The sizes determined for fragments observed in lanes 1-4 are listed in Table 6.1.

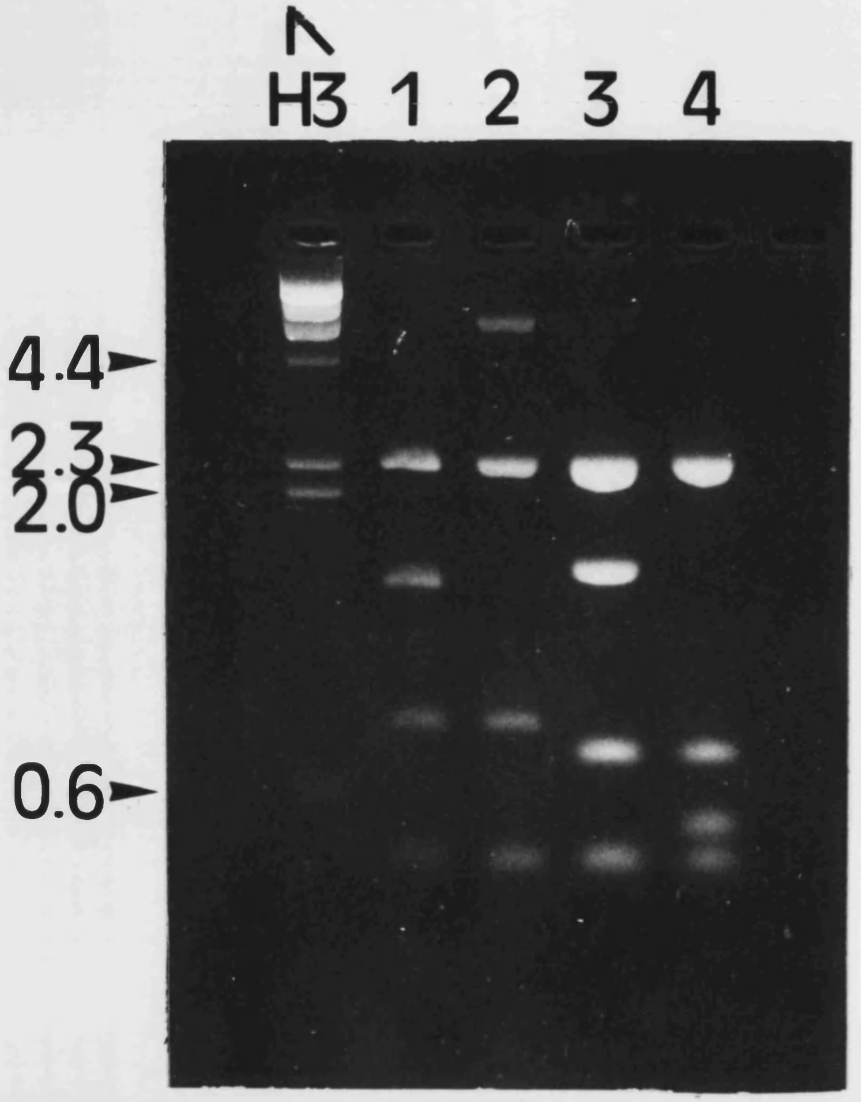

λ
H3
1
2
3
4
4.4
2.3
2.0
0.6

Table 6.1: Analysis of Tp.acidophilum citrate synthase constructs with PvuII

Construct	Fragment(kb) Expected	Determined
pTaCS18	2.37	2.4
	~1.4	1.5
	0.76	0.8
	0.41	0.4
pCSEH18	2.37	2.4
	0.76	0.8
	0.41	0.4*
	0.40	0.4*
pTaCS19	2.37	2.4
	~1.5	1.5
	0.67	0.7
	0.41	0.4
pCSEH19	2.37	2.4
	0.67	0.7
	0.50	0.5
	0.41	0.4

See Fig. 6.2 for experimental details.

* these fragments were indistinguable on the 0.6% (w/v) agarose gel.

Figure 6.3: Sequences of upstream regions of pTaCS18, pTaCS19, pCSEH18, pCSEH19 and pCSSS

The reading frames indicated are those obtained by initiation of translation from the lac ribosome-binding site. The nucleotides underlined with asterisks represent the putative ribosome-binding site for the Tp.acidophilum citrate synthase gene. The nucleotides underlined are those representing (or formerly involved in representing) a restriction site for the enzyme(s) indicated*. The nucleotides in lower case are those involved in coding for the Tp.acidophilum citrate synthase.

* HII-EV: former HincII site ligated to former EcoRV site; S: former SphI site.

(a) pTaCS19

```
     T   M   I   T   P   S   L   H   A   C   R   S   T   L   E   D
ATG ACC ATG ATT ACG CCA AGC TTG CAT GCC TGC AGG TCG ACT CTA GAG GAT CC
lacZ ->                          SphI          HincII          BamHI

--approximately 1-kb---AAAAGATATCATATATGTAGAGGTGTATTAATGccag---etc.
                        EcoRV          ****
```

(b) pCSEH19

```
     T   M   I   T   P   S   L   H   A   C   R   S   S   Y
ATG ACC ATG ATT ACG CCA AGC TTG CAT GCC TGC AGG TCA TCA TAT
lacZ ->                          SphI          HII-EV

 M  stop
ATG TAG AGG TGT ATT AAT Gcc aga aac tg aa ga aat ta -----etc.
       *** *
```

(c) pCSSS

```
     T   M   I   T   P   S   L   P   A   G   H   H   I
ATG ACC ATG ATT ACG CCA AGC TTG CCT GCA GGT CAT CAT ATA
lacZ ->                          S           HII-EV

 C   R   G   V   L   M   P   E   T   E   E
TGT AGA GGT GTA TTA ATG cca gaa act gaa gaa ------etc.
      * ***
```

(d) pTaCS18

```
CCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCC-----approximately 1-kb--
        SphI        HincII        BamHI

--AAAAGATATCATATATGTAGAGGTGTATTAATGccagaaactgaagaa----etc. <- lacZ
      EcoRV           ****
```

(e) pCSEH18

```
CCAAGCTTGCATGCCTGCAGGTCATCATATATGTAGAGGTGTATTAATGccaga--etc. <- lacZ
        SphI        HII-EV          ****
```

depicted in Fig. 6.1.(a) and (b).

A unique EcoRV site, located at position "-26" of the citrate synthase gene and a unique HincII site, located in the original pUC polylinker, were present in pTaCS18 and pTaCS19. The constructs were digested with these two enzymes. The resulting linear vector DNA was separated from the excised 1-kb fragments on a 0.8% (w/v) LMP agarose gel and purified by the freeze-squeeze method. The vectors were religated (both EcoRV and HincII cut to give blunt-ended DNA) and transformed into competent E.coli TG1 cells; constructs, designated pCSEH18 and pCSEH19, were purified from the resulting clones. PvuII-digestion of pCSEH18 and pCSEH19 (Fig. 6.2, Table 6.1) and partial sequencing of them (Fig. 6.3.b & e) confirmed that the desired 1-kb fragments had been removed and the vectors had religated in the predicted manner.

6.2.3. Construction of plasmid pCSSS

In order to allow translation of the gene from the lac ribosome-binding site, the citrate synthase gene in pCSEH19 was brought into the correct reading frame by the scheme outlined below and depicted in Fig. 6.1.c.

A unique site for SphI, located at position "-38" was present on pCSEH19; the plasmid was digested with this enzyme. The linear vector was then treated with S1 nuclease in order to remove the four nucleotide overhang produced as a result of the action of SphI. The vector DNA (now blunt-ended) was religated and transformed into competent TG1 cells; the construct, designated pCSSS, was purified from a resulting clone. A partial sequencing of pCSSS confirmed that the citrate synthase gene was in the same reading frame as the start of the lacZ gene (Fig. 6.3.c).

6.2.4. Comparison of expression from pTaCS18, pTaCS19, pCSEH18, pCSEH19 and pCSSS

E.coli strains containing pUC19, pTaCS18, pTaCS19, pCSEH18, pCSEH19 and pCSSS were grown as described in Section 2.2.27. Citrate synthase activities from crude cell-free extracts (from three separate growth experiments) are listed in Table 6.2.

Samples of cell-free extracts from pUC19 and pTaCS19 clones (assay 1) were run on an SDS-polyacrylamide gel (Fig. 6.4). An additional protein (M_r = 44 000 ± 2 000) was observed in the pTaCS19 cell-free extract.

Samples of cell-free extract from pUC19, pTaCS18, pCSEH18, pCSEH19 and pCSSS clones (assay 3) were analysed on an SDS polyacrylamide gel (Fig. 6.5). An additional protein (M_r = 44 000 ± 2 000) was observed in cell-free extracts from the pTaCS18, pCSEH19 and pCSSS clones but not in that of the pCSEH18 clone. A second additional protein (M_r = 47 000 ± 2 000) was observed in the pCSSS cell-free extract.

6.3. Discussion

6.3.1. Evidence for expression of Tp.acidophilum citrate synthase

The cell-free extract of a pUC19 clone showed a very low background level of citrate synthase activity. This demonstrated that carrying out the assays at 55°C ensured that the E.coli citrate synthase was essentially inactivated. So, activity observed for E.coli strains carrying constructs of the Tp.acidophilum citrate synthase indicated that functional Tp.acidophilum citrate synthase had been expressed.

Further evidence that Tp.acidophilum citrate synthase was being produced was obtained by SDS-PAGE analysis of the clones. When cell-free extracts from strains carrying pTaCS18, pTaCS19, pCSEH19 and pCSSS were compared with that from a strain carrying pUC19 by SDS-PAGE (Figures 6.4 & 6.5), an additional band

Table 6.2: Activities of Thermoplasma acidophilum citrate synthase in crude extracts of E.coli strains carrying pUC, pTaCS18, pTaCS19, pCSEH18, pCSEH19 and pCSSS

		Specific activity (units/mg protein)*		
Construct	IPTG (mM)	Assay 1	Assay 2	Assay 3
pUC19	0	0.3	nd	0.01
pUC19	10	0.3	0.1	nd
pTaCS18	0	0.7	0.3	0.6
pTaCS18	10	0.6	0.3	nd
pTaCS19	0	0.5	0.3	nd
pTaCS19	10	0.4	0.3	nd
pCSEH18	0	nd	0.1	0.05
pCSEH18	10	nd	0.1	nd
pCSEH19	0	nd	0.8	4.5
pCSEH19	10	nd	0.9	nd
pCSSS	0	nd	0.5	4.4
pCSSS	10	nd	0.4	nd

Assays 1, 2 and 3 refer to results obtained from three different growth experiments.

* A unit of citrate synthase activity is defined as μmol CoASH produced/min.

nd = not determined.

Figure 6.4: Separation of cell-extracts of E.coli carrying pUC19 (track 2) and pTaCS19 (track 3) by SDS-PAGE

Molecular weights of the protein standards (track 1) are indicated in the left-hand margin. The arrow in the right-hand margin indicates an additional band which corresponds approximately to the size predicted for Tp.acidophilum citrate synthase (i.e. 42 942) by sequencing of the gene (Chapter 3).

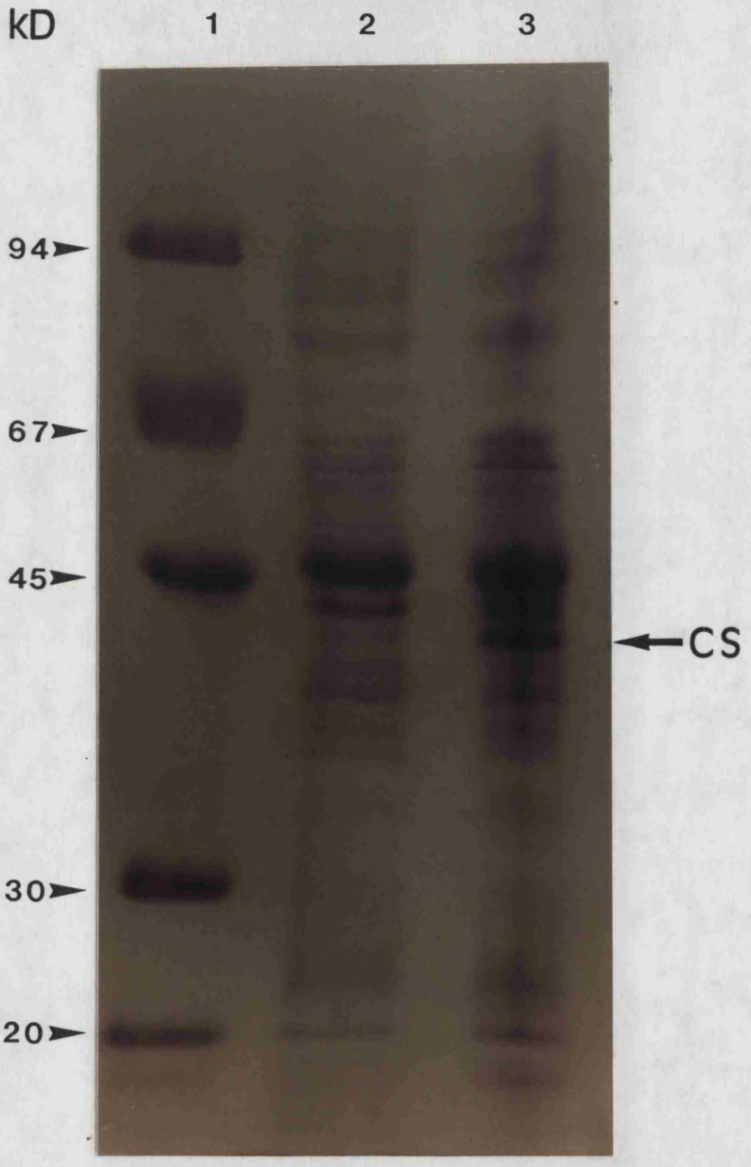
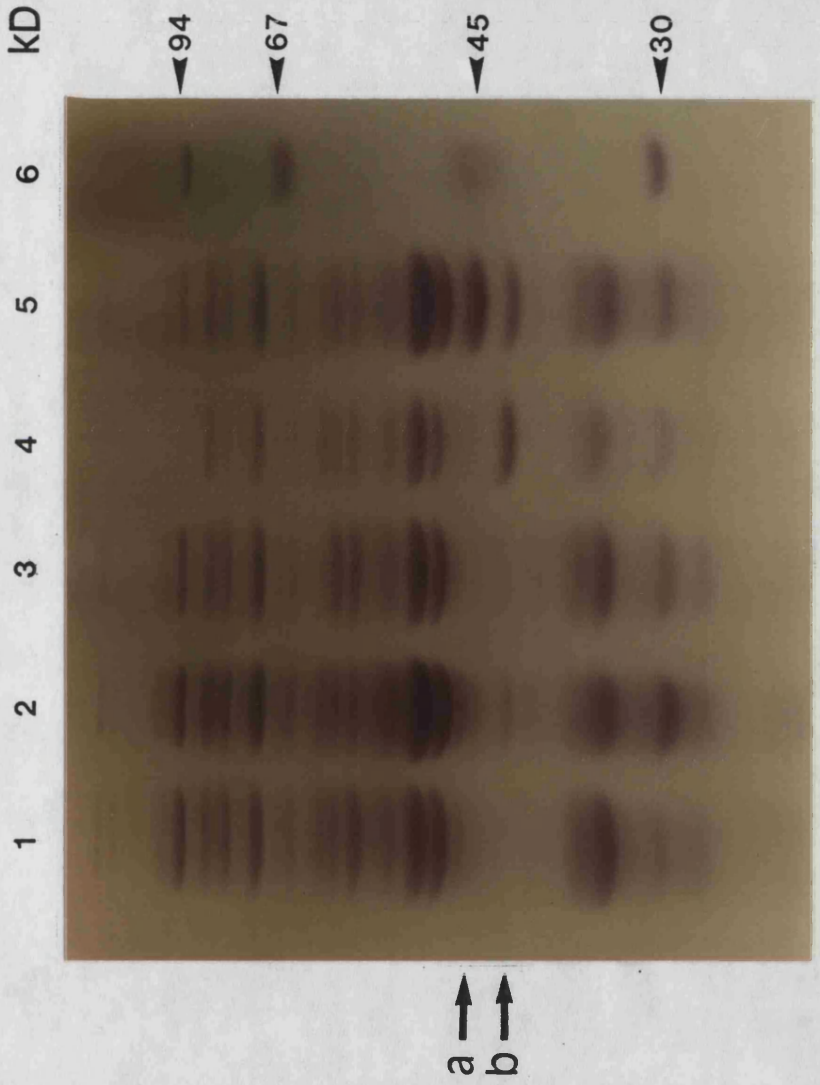

kD
1
2
3
94
67
45
30
20
CS

Figure 6.5: Separation of cell-extracts of E.coli carrying pUC19 (track 1), pTaCS18 (track 2), pCSEH18 (track 3), pCSEH19 (track 4) and pCSSS (track 5) by SDS-PAGE

Molecular weights of protein standards (track 6) are indicated in the right-hand margin. Arrows in the left-hand margin indicate additional bands due to expression of Tp.acidophilum citrate synthase (b) and a ß-galactosidase-citrate synthase fusion protein (a).


KD
94
67
45
30
6
5
4
3
2
1
a
b

corresponding to protein of M_r 44 000 ± 2 000 was observed. This size agrees with that predicted for Tp.acidophilum citrate synthase from its amino acid composition (M_r = 42 942) (Chapter 3) and that observed for the pure protein isolated from Tp.acidophilum (M_r = 43 000 ± 2 000) [Smith et al. (1987)]. A second additional band corresponding to protein of M_r 47 000 ± 2 000 was observed in cell-extract from the pCSSS clone. This size agrees with that predicted for the expected ß-galactosidase-citrate synthase fusion protein (M_r = 44 977) (Fig. 6.3.c).

6.3.2. Level of Tp.acidophilum citrate synthase expression in E.coli

The activity of citrate synthase from crude extracts of Tp.acidophilum cells has been found to be 0.2 units per mg protein [Smith et al. (1987)], whereas as much as 4.5 units per mg protein were observed, in this study, for Tp.acidophilum citrate synthase expressed in E.coli. This indicates that the enzyme was being expressed in E.coli at a level up to 20-times higher than that found in the native cell. The activity of pure Tp.acidophilum citrate synthase has been found to be approximately 41 units per mg protein [Smith et al. (1987)]; thus, the 4.5 units per mg protein of citrate synthase, obtained from the construct pCSEH19, indicates that the Tp.acidophilum citrate synthase was approximately 10% of total cell protein in the strain carrying this plasmid.

6.3.3. Features of the Tp.acidophilum citrate synthase expression system

Generally, it is preferable that expression of a foreign gene is inducible since this allows initial unchecked growth of the host - if desired. IPTG is normally required to initiate transcription from the lac promoter of pUC. However, in this study, citrate synthase was produced, even in the absence of IPTG, from

some constructs that require the lac promoter (Table 6.2, i.e. pCSEH19 and pCSSS - see Section 6.3.4). Similar findings have been obtained by Bright (1990) for expression of the Tp.acidophilum glucose dehydrogenase gene in TG1 cells. An additional finding, where TG1 clones of pUC turned blue when grown in the presence of X-gal but in the absence of IPTG, suggests that the lacZ gene is "switched on" in the TG1 cells even in the absence of the artificial inducer. So E.coli TG1 cells, it seems, may lack adequate repression and/or may possess a natural inducer of the lac promoter. This lack of inducibility of constructs in E.coli TG1 cells, however, was not regarded as a problem for the production of Tp.acidophilum citrate synthase, since it still appeared to be produced at reasonable levels in these cells (i.e. up to 10% of total cell protein).

The codon usage observed in the Tp.acidophilum citrate gene is different from that found in E.coli genes (discussed in Chapter 4). However, this does not seem to have prevented production of the Tp.acidophilum protein in relatively high amounts.

It is clear from the results of the three separate assay experiments, listed in Table 6.2, that levels of expression from the same construct varied enormously. This was probably due to inconsistencies in growth conditions (possibly the duration of growth) and/or variations in the sonicating process. Further work is necessary in order to optimise the growth conditions required for optimal expression.

6.3.4 Comparison of expression from pTaCS18, pTaCS19, pCSEH18, pCSEH19 and pCSSS

Even though expression levels varied between separate growth experiments, it was still possible to observe consistent variations in expression between different constructs within each experiment. Conclusions regarding these variations will now be discussed.

Citrate synthase was expressed at similar levels from both pTaCS18 and pTaCS19, even though the gene in the former construct is in the opposite orientation to that required for initiation of transcription from the lac promoter. This result implies that a region of Tp.acidophilum DNA, upstream of the citrate synthase gene may be recognised by the E.coli RNA polymerase.

Whereas citrate synthase was expressed from pCSEH19 at high levels, no expression was observed from pCSEH18. This lack of expression from pCSEH18 suggests that the region of Tp.acidophilum DNA, allowing transcription in pTaCS18 and pTaCS19, is no longer present in pCSEH18 (and pCSEH19). The variation in expression from pCSEH18 and pCSEH19 can be explained by the difference in orientation of the Tp.acidophilum gene in relation to the lac promoter. In pCSEH18, the gene is in the wrong orientation, whereas in pCSEH19 it is in the correct orientation, to allow its transcription to proceed from the lac promoter.

Citrate synthase was expressed at a higher level from pCSEH19 than from pTaCS19, indicating that the lac promoter (used in pCSEH19) is stronger than the Tp.acidophilum "promoter" (used in pTaCS19). The intervening 1-kb non-coding region of Tp.acidophilum DNA (present in pTaCS19 but not in pCSEH19) seems to prevent the more efficient transcription of the gene from the lac promoter in pTaCS19.

The observed expression from pTaCS18 must be the result of initiation of translation from the ribosome-binding site of the citrate synthase gene, since the gene in this construct is in the wrong orientation in relation to the lac ribosome-binding site. In pTaCS19 and pCSEH19, translation is probably also due to initiation from the citrate synthase gene ribosome-binding site. In pTaCS19, any translation initiated from the lac ribosome-binding site is likely to terminate within the intervening 1-kb of DNA prior to the start of the gene. Whereas, in pCSEH19, translation from the lac ribosome-

binding site is not in frame with that required for the citrate synthase gene (Fig. 6.3.b) and, in any case, probably terminates at the "TAG" termination codon found immediately in front of the gene's own ribosome-binding site (Fig. 6.3.b).

Citrate synthase was expressed from pCSSS at a level equivalent to that observed from pCSEH19 (Fig. 6.5). In addition to the citrate synthase, a ß-galactosidase-citrate synthase fusion protein appeared to be produced from pCSSS, presumably as a result of initiation of translation from the lac ribosome-binding site (Fig. 6.3.c). Even though this fusion protein was expressed at high levels (as judged by PAGE, Fig. 6.5), the observed citrate synthase activity from pCSSS was still equivalent to that observed from pCSEH19 (Table 6.2), suggesting that the fusion protein is essentially inactive.

In conclusion, among the constructs available, pCSEH19, in which the Tp.acidophilum gene is transcribed from the lac promoter and translated from its own ribosome-binding site, gave the optimum expression of a functional citrate synthase. This construct was used for the purification studies discussed in Chapter 7.

CHAPTER 7: PURIFICATION OF Thermoplasma acidophilum CITRATE SYNTHASE EXPRESSED IN Escherichia coli

7.1. Introduction

This chapter describes experiments designed to purify Tp.acidophilum citrate synthase expressed in E.coli.

Heat stability studies on Tp.acidophilum citrate synthase carried out by Grossebüter & Görisch (1985) demonstrated that the enzyme retains 70% of its activity when incubated at 70°C for 10 min. This thermostability of Tp.acidophilum citrate synthase suggested the use of a heat step as the basis for purification of the enzyme from E.coli, since thermophilic proteins expressed in mesophilic hosts favour a simple and effective purification from the heat-unstable host proteins [Fabry et al. (1988); Tiboni et al. (1989)].

Chromatofocussing, the most successful step in the purification of the citrate synthase from Tp.acidophilum [Smith et al. (1987)], was selected as a possible step for the purification of the enzyme expressed in E.coli.

7.2. Results

7.2.1. Heat purification of Tp.acidophilum citrate synthase expressed in E.coli

Aliquots (1 ml) of cell-free extract of the pCSEH19 clone (prepared as described in Chapter 2) were incubated at 65°C, 75°C and 85°C for 15 min. Separation on an SDS-polyacrylamide gel (Fig. 7.1) and comparison of specific activities (listed in Table 7.1.) suggested that 65°C was the optimum temperature for heat purification of the Tp.acidophilum citrate synthase from E.coli (see Section 7.3).

Figure 7.1: SDS-polyacrylamide gel showing heat purification of E.coli-produced Tp.acidophilum citrate synthase

From left to right: standard protein markers; (1) cell-free extract; (2) after heat precipitation at 65°C; (3) after heat precipitation at 75°C; and (4) after heat precipitation at 85°C. Molecular weights of the protein standards are indicated in the left-hand margin.

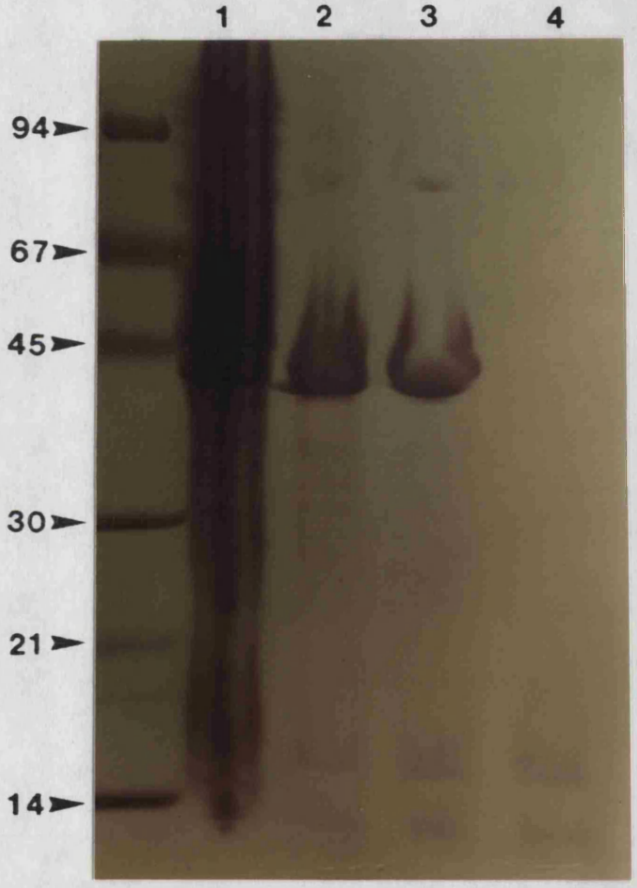

1
2
3
4
94
67
45
30
21
14

Table 7.1: Citrate synthase activities in heat-treated samples of E.coli TG1 containing pCSEH19

Sample	Volume (ml)	Total enzyme (units)	Specific activity (units/mg protein)	Recovery (%)	Overall purification (-fold)
Untreated	1.0	0.7	4	100	-
65°C	0.9	0.6	30	86	7.5
75°C	0.9	0.6	31	86	7.8
85°C	0.9	3×10^{-4}	0.03	0	-

7.2.2. Chromatofocussing of Tp.acidophilum citrate synthase expressed in E.coli

Cell-free extract was prepared from a 250 ml culture of the pCSEH19 clone and incubated at 65°C for 15 min. The sample was dialysed against 25 mM diethanolamine-HCl (pH 9.5) and subjected to chromatofocussing on an FPLC mono P column. The Tp.acidophilum enzyme was found to elute from the column over a pH range of pH 8.7 - 8.0. An elution profile of the citrate synthase is shown in Fig. 7.2.

Three fractions contained the majority of the citrate synthase activity. These were pooled and filtered through a polysulfone filter. The sample was resuspended in a 20 mM Tris-HCl buffer (pH 8.0) containing 1 mM EDTA to a final concentration of 10 $mg.ml^{-1}$ protein.

The sample gave a single major band on an SDS-polyacrylamide gel (Fig. 7.3), which corresponded to a protein M_r 42 000 (± 2 000). A gel scan (Fig. 7.4) obtained on an LKB Ultrascan laser densitometer indicated that the sample was > 95% pure.

A summary of purification results obtained by the above protocol is presented in Table 7.2.

7.2.3. Estimation of molecular size of Tp.acidophilum citrate synthase

Data collected on an Oros Instruments Molecular Size Detector, as the result of analysis of Tp.acidophilum citrate synthase (purified as described in Section 7.2.2), are listed in Table 7.3. The average molecular mass of the citrate synthase was estimated to be 77 000 (± 2 000). This indicated that native Tp.acidophilum citrate synthase is dimeric (n = 1.8), the subunit molecular weight having been determined to be 42 942 by sequencing of the gene (Chapter 3).

Figure 7.2: FPLC MonoP elution profile for a heat-treated cell extract of E.coli strain expressing Tp.acidophilum citrate synthase

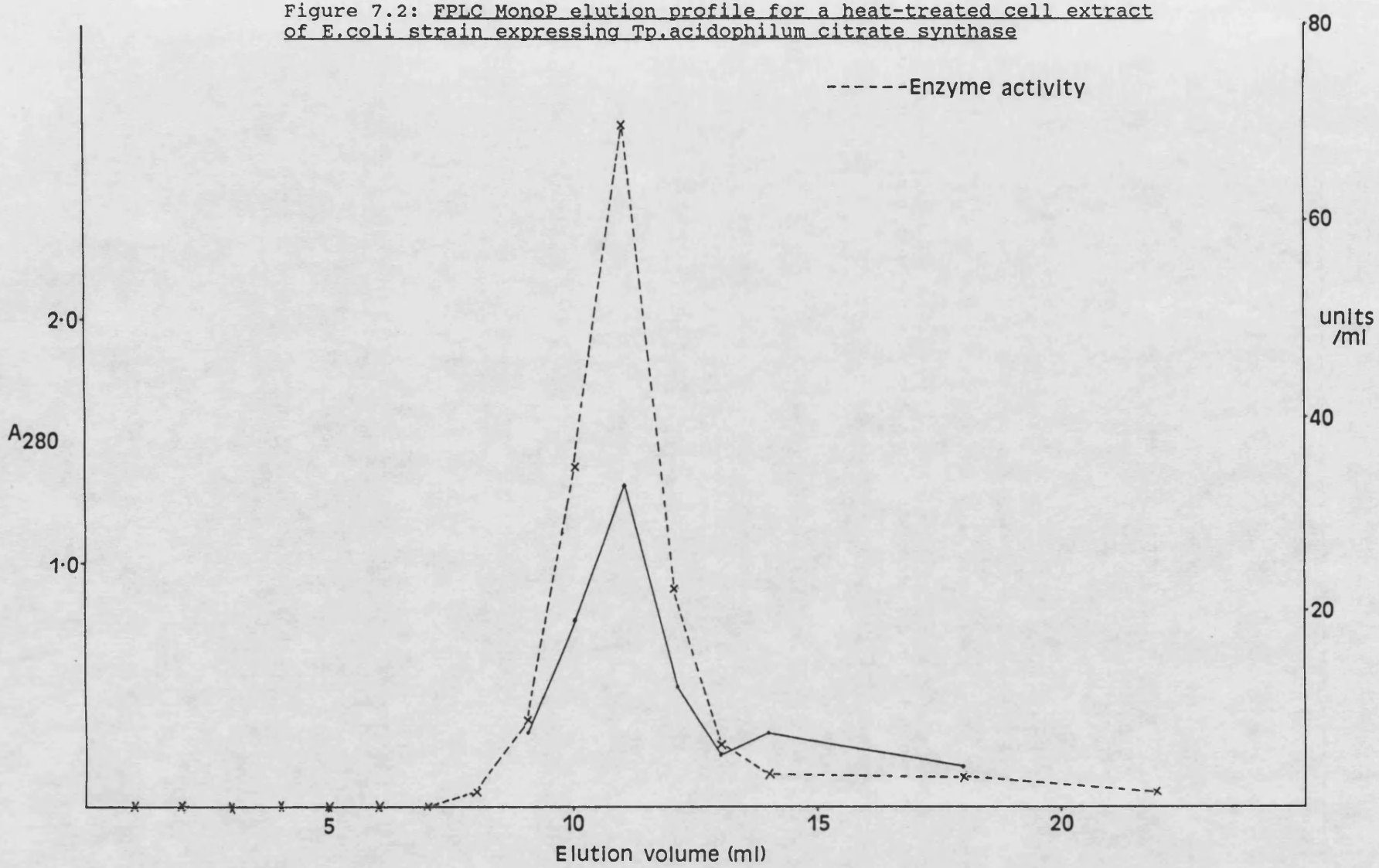


Figure 7.3: SDS-PAGE analysis of fractions from the purification of E.coli-produced Tp.acidophilum citrate synthase

Lanes: (1) cell-free extract; (2 & 3) after heat precipitation at 65°C; (4) after chromatofocussing (FPLC Mono P).

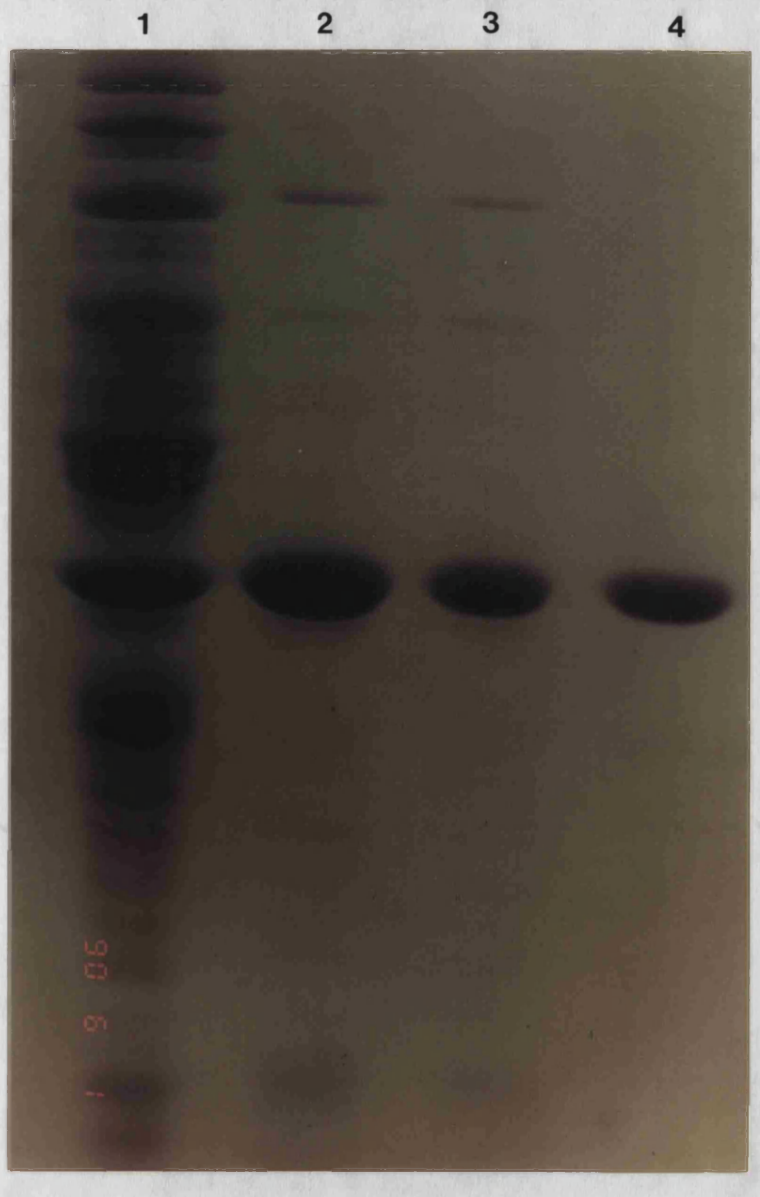

1
2
3
4

Figure 7.4: Linear scan of lane (4) of gel shown in Fig. 7.3 as produced by a LKB ultrascan densitometer

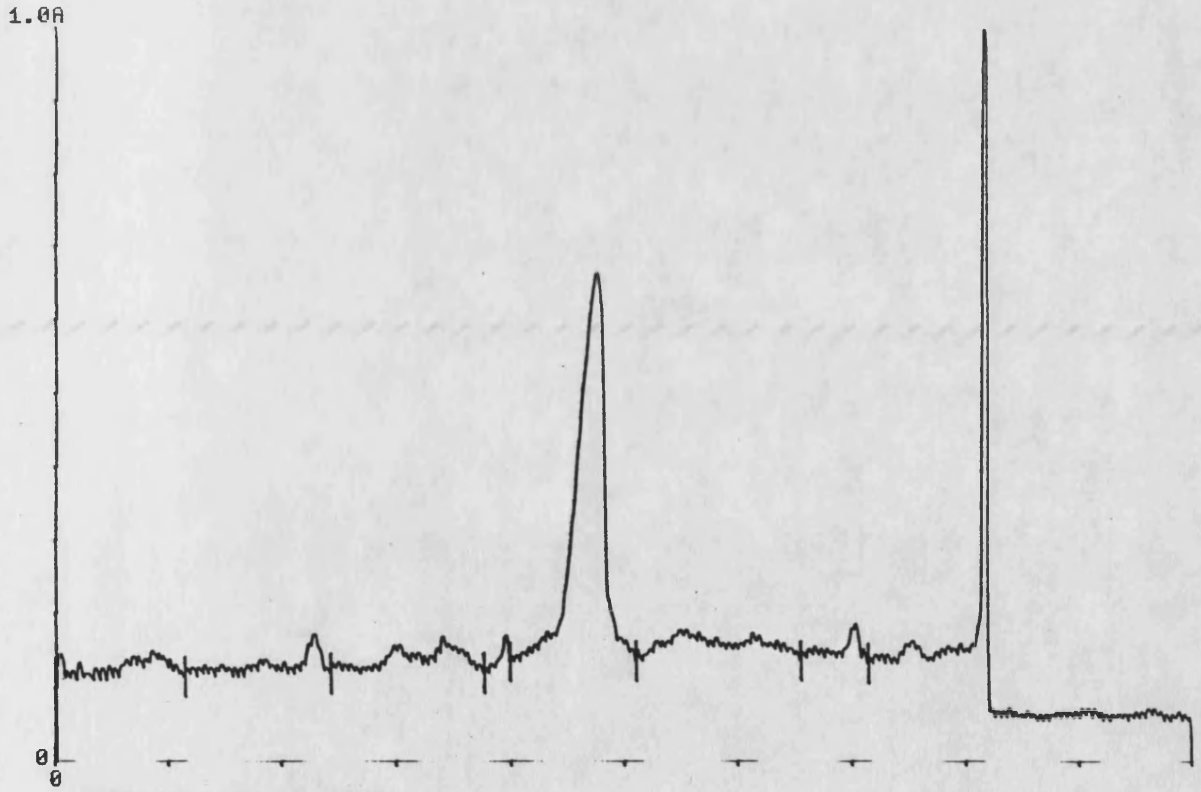
The main peak corresponds to the Tp.acidophilum citrate synthase and represents > 95% of protein present.

```
 LINEAR SCAN PARAMETERS:
------------------------
 NO. OF SCANS: 1
 SCAN LENGTH: 100
 APERTURE WIDTH: 0.1
 MODE: "%"= 0, ABSORBANCE= 1: 1
 MAX. ABSORBANCE: 1.0
 GRAPHIC OUTPUT: (INVERT=1,NORMAL=0) 0
 BACKGROUND CORRECTION: (NO=0,VALLEY TO VALLEY=1,RUBBER BAND METHOD=2) 0
 AUTO ZERO: (YES=1,NO=0) 0
 PEAK RESULTS: (YES=1,NO=0) 1
 PEAK/TROUGH THRESHOLDS: HEIGHT= 8 WIDTH=  NOISE=
 POST SCAN EDIT: (NO=0,AFTER GRAPH PLOT=1,AFTER RESULTS PRINTOUT=2) 0

 LINEAR SCAN PARAMETERS:
------------------------
 NO. OF SCANS:
 LINEAR SCAN RUNNING...
------------------------

   SCAN NUMBER= 2

   SCAN LENGTH= 100 MM. APERTURE WIDTH= 0.1 MM.
```




```
   LENGTH OF X-AXIS= 100 MM.

   TOT. INTEGRAL= 72258
```

PEAK	POSITION	HEIGHT	REL%	INTEGRAL	PEAK	POSITION	HEIGHT	REL%	INTEGRAL
1	8.7	38	10.26	7414	2	23.0	44	11.66	8432
3	34.2	42	12.9	9324	4	39.8	44	2.07	1500
5	47.8	169	20.38	14730	6	55.1	46	15.95	11528
7	70.5	47	6.13	4432	8	81.7	255	20.61	14898

Table 7.2: Purification of Tp.acidophilum citrate synthase from E coli TG1 containing pCSEH19

Purification step	Volume (ml)	Total enzyme (units)	protein (mg)	Specific activity (units/mg)	Recovery (%)	Overall purification (-fold)
Cell extract	4.0	320	60	5	100	-
Heat treatment (65°C)	2.8	241	10	24	75	4.8
Chromatofocussing (Mono P)	3.0	148	3.5	42	46	8.4

Table 7.3: Molecular sizes of Tp.acidophilum citrate synthase as determined on an Oros Instruments Model 801 Molecular Size Detector

The citrate synthase was purified as described in Section 7.2.2. An average of the molecular mass estimations is quoted in the text.

Time (mm.ss)	Amp.	Diffn Coeff	Radius (nm)	Est. MW	Conc mg/ml	Polyd. (nm)	Purity (%)	UV (mV)	Temp °C	Count Rate	Meas Code	Anal. Error	Base Line	SOS before	SOS after
0000.00	0.682	646	3.8	75k	0.51	1.348		0.09	25.8	47	I	11	1.002	2.772	2.430
0000.40	0.693	658	3.7	71k	0.53	1.088		0.10	25.8	47			1.001	4.491	4.248
0000.61	0.683	631	3.9	80k	0.49	1.539		0.10	26.0	48	SI	11	1.002	6.281	5.521
0000.83	-----	-----	-----	-----	-----	-----		0.10	25.8	72		1	-----	-----	-----
0000.83	0.681	617	4.0	83k	0.48	1.972		0.10	25.8	49	BSI	11	1.006	11.283	9.937
0001.23	0.685	654	3.8	73k	0.52	1.486		0.10	26.0	47	SI	11	1.002	6.716	6.017
0001.45	0.690	650	3.8	74k	0.52	1.488		0.10	25.8	47	SI	11	1.000	8.301	7.446
0001.66	0.686	650	3.8	74k	0.52	1.494		0.09	26.0	48	SI	11	1.003	7.215	6.425
0001.88	0.683	659	3.7	71k	0.54	1.414		0.11	25.9	47	SI	11	1.002	6.223	5.789
0002.10	0.671	668	3.7	70k	0.56	1.454		0.10	26.0	48	SI	11	1.002	5.476	4.968
0002.30	0.681	660	3.7	72k	0.53	1.477		0.09	26.1	47	SI	11	1.000	5.751	5.088
0002.53	0.667	670	3.7	69k	0.55	1.071		0.09	25.9	47			1.004	2.103	1.914
0002.73	0.672	627	3.9	80k	0.50	1.542		0.09	25.8	49	SI	11	1.003	5.605	5.132
0002.95	0.674	600	4.1	90k	0.47	1.615		0.09	25.9	51	BSI	11	1.008	6.209	5.406
0003.36	0.675	582	4.2	97k	0.45	2.098		0.09	25.9	53	BSI	11	1.009	12.009	10.602
0003.56	0.670	628	3.9	81k	0.49	1.547		0.09	26.0	49	I	11	1.004	4.583	3.973
0003.98	0.682	661	3.7	71k	0.54	1.321		0.09	25.9	48	I	11	1.001	4.276	3.955
0004.18	0.661	661	3.7	71k	0.55	1.467		0.11	25.9	48	SI	11	1.001	9.818	9.818
0004.40	0.663	626	3.9	81k	0.48	1.547		0.11	25.9	48	SI	11	1.001	8.282	7.587
0004.61	0.661	656	3.8	74k	0.53	1.493		0.09	26.3	49	I	11	1.003	4.644	4.172
0004.81	0.664	652	3.8	74k	0.54	1.487		0.11	26.0	49	SI	11	1.003	6.065	5.615
0005.03	0.661	618	4.0	83k	0.49	1.914		0.09	25.9	51	BSI	11	1.007	8.791	7.748
0005.25	0.663	663	3.7	70k	0.56	1.458		0.09	25.8	48	I	11	1.002	3.856	3.397
0005.66	-----	-----	-----	-----	-----	-----		0.10	25.8	55		1	-----	-----	-----

7.3. Discussion

The thermostability of the Tp.acidophilum citrate synthase provided a basis for its rapid and simple purification from E.coli proteins. A comparison of temperatures indicated that 65°C was optimal for the partial purification of the citrate synthase by heat treatment. At 75°C, no increase in purity over incubation at 65°C was obtained and, at 85°C, the enzyme was denatured (Table 7.1, Fig. 7.1).

Chromatofocussing proved a useful step in purification of the Tp.acidophilum citrate synthase from heat-treated E.coli cell-free extract. The citrate synthase was one of the first proteins to detach from the column on application of the polybuffer (i.e. at pH 8.0 - 8.7), indicating that the enzyme has a relatively high isoelectric point. Following chromatofocussing, the specific activity of the citrate synthase (42 units/mg protein) was comparable to that obtained for the enzyme purified from Tp.acidophilum (41 units/mg protein) [Smith et al. (1987)]. This indicates that purity of the enzyme had been achieved to a similar level.

The molecular mass of the native Tp.acidophilum citrate synthase was estimated to be 77 000 (± 2 000) using a molecular size detector. This differs slightly from the value of 85 000 (± 2 000) determined by gel filtration [Danson et al. (1985); Smith et al. (1987)]. The molecular size detector measures the radius of a molecule, then, making the assumption that the molecule is spherical, it estimates the molecular mass. However, even a globular protein is not necessarily spherical, and so the size detector's estimation of molecular mass is unlikely to be accurate. Moreover with the size detector, as with gel filtration, the molecular mass of a compact molecule will tend to be underestimated, and the molecular mass of a "floppy" molecule overestimated. So, the figure determined for the Tp.acidophilum citrate synthase, using the molecular size detector, although a

good confirmation of the enzyme being dimeric ($n \simeq 2.0$), cannot be taken to be an accurate estimate of its molecular mass.

Tp.acidophilum citrate synthase, purified as described in Section 7.2.2, was judged pure enough (by SDS-PAGE) to initiate crystallisation trials. These are now in progress. The ultimate aim of this work is to obtain X-ray data for determination of the 3D-structure of the protein.

CHAPTER 8: GENERAL DISCUSSION AND SUGGESTIONS FOR FURTHER WORK

8.1. Cloning strategy

The cloning of protein-encoding genes (including archaebacterial genes), using a variety of different approaches, is now well-documented in the literature. The construction of "mini-libraries", use of oligonucleotides, and a directional cloning approach were important aspects of the cloning strategy used to clone the Tp.acidophilum citrate synthase gene.

Oligonucleotides based on sequence information were used to identify the Tp.acidophilum citrate synthase gene. This method was chosen as it did not require functional expression of the enzyme. The design of the oligonucleotide proved crucial to the successful identification of the citrate synthase gene. Whereas a long, non-redundant oligonucleotide was effective in identifying the citrate synthase gene, a short, redundant oligonucleotide did not prove specific enough (Chapter 3).

8.2. Sequence information

The derived amino acid sequence of the Tp.acidophilum citrate synthase was compared by multiple-sequence alignment analysis with eubacterial and eukaryotic citrate synthase sequences. The program used had been developed specifically for citrate synthase sequences, and takes into account the regions of secondary structure known for the pig enzyme [Henneke et al. (1989)]. Stringency of the alignment was enhanced by including all available eight citrate synthase sequences in the comparison.

From the citrate synthase sequence alignment, it appears that although there is a high degree of homology between citrate synthases within a kingdom, there is very little homology between citrate synthases of different kingdoms. Nevertheless, the majority of catalytically

important residues and those involved in binding substrate are conserved in the Tp.acidophilum enzyme. This lack of sequence similarity along with conservation of functionally important residues has been reported for comparisons between other archaebacterial and non archaebacterial proteins [for examples see Beckler & Reeve (1986); Lechner & Böck (1987); Cubellis et al. (1989); Hensel et al. (1989); Honka et al. (1990); Sanangelantoni et al. (1990)].

On the basis of rRNA sequence analysis, the Thermoplasma genus groups with the methanogens and halophiles, and not as might be expected with the sulphur-dependent thermophiles [Yang et al. (1985)]. Given that citrate synthase is present in all three archaebacterial phenotypes [Danson et al. (1985)], further sequencing of this enzyme from the halophiles, methanogens and other thermophiles is a priority. Such data will not only contribute to the debate over the relationship between the proposed three primary kingdoms, but may also throw light on the unusual phylogenetic position of Tp.acidophilum within the archaebacteria.

An analysis of amino acid changes occurring among the Tp.acidophilum, pig and E.coli citrate synthases was carried out in order to identify possible thermophilic features of the Tp.acidophilum protein (Chapter 5). However, the phylogenetic differences between these three organisms proved too great to allow any conclusions regarding thermostability to be drawn. A more conclusive analysis may be possible once additional archaebacterial citrate synthase sequences are obtained.

The sequence of the Tp.acidophilum citrate synthase was analysed in order to predict its secondary structure but, as judged from results for the pig enzyme, the prediction is unreliable (Chapter 5). A more productive structural comparison may be achieved once the 3D-structure of the Tp.acidophilum enzyme is determined.

In the absence of direct experimental proof, no specific conclusions regarding the regulatory regions of the Tp.acidophilum citrate synthase gene can be drawn. Nevertheless, on the basis of comparison with published archaebacterial consensus sequences [Zillig et al. (1988); Thomm & Wich (1988); Brown et al. (1989)], putative regulatory regions have been proposed (Chapter 4), and these should now be subjected to closer scrutiny.

8.3. Expression of the Tp.acidophilum citrate synthase gene

The problem of poor cell yields encountered when growing archaebacteria [Smith (1989)] can be overcome by expressing archaebacterial genes in E.coli. Therefore, one of the aims of this project was to express the Tp.acidophilum citrate synthase in E.coli (see Chapter 6). Activity experiments and PAGE analysis demonstrated that an initial construct containing the Tp.acidophilum citrate synthase gene, pTaCS19, was producing catalytically-active citrate synthase in E.coli. However, a construct, pCSEH19, generated from pTaCS19, showed greater levels of expression. Judging from citrate synthase activity, up to 10% of total protein content of the E.coli strain carrying pCSEH19 was Tp.acidophilum citrate synthase.

Inconsistencies in the levels of expression of Tp.acidophilum citrate synthase, due to variations in growth and sonication conditions, were observed. These need to be eliminated in order to allow optimal expression to be achieved.

8.4. Purification of Tp.acidophilum citrate synthase

The expression of a thermophilic protein in a mesophilic host allows a simple and rapid purification of the protein. This was amply illustrated in the purification of Tp.acidophilum citrate synthase from the mesophilic E.coli proteins (Chapter 7). Most of the host

proteins were removed by a simple heat-treatment (at 65°C) of the cell-extract followed by centrifugation. The citrate synthase was then purified to homogeneity by a chromatofocussing of the heat-treated sample.

Future long-term aims of the work include:
(1) obtaining crystals of the Tp.acidophilum citrate synthase for determination of its 3D-structure; and
(2) site-directed mutagenesis and kinetic analysis of the protein.

Such work, it is hoped, will help elucidate catalytic and thermostable features of the Tp.acidophilum citrate synthase.

REFERENCES

Achenbach-Richter, L., Gupta, R., Stetter, K.O. & Woese, C.R. (1987) System. Appl. Microbiol. 9, 34-39.

Ahern, T. J., Casal, J.I., Petsko, G.A. & Klibanov, A.M. (1987) Proc. Natl. Acad. Sci. USA 84, 675-679.

Alber, T. & Mathews, B,W. (1987) Meth. Enzym. 154, 511-533.

Amils, R., Ramírez, L., Sanz, J.L., Marín, I., Pisabarro, A.G. & Ureña, D. (1989) Can. J. Microbiol. 35, 141-147.

Anderson, D.H. & Duckworth, H.W. (1988) J. Biol. Chem. 263, 2163-2169.

Argos, P., Rossmann, M.G., Grau, U.M., Zuber, H., Frank, G. & Tratschin, J. D. (1979) Biochemistry 18, 5698-5703.

Beckler, G.S. & Reeve, J.N. (1986) Mol. Gen. Genet. 204, 133-140.

Bhayana, V. & Duckworth, H.W. (1984) Biochemistry 23, 2900-2905.

Bloxham, D.P., Parmlee, D.C., Kumar, S., Wade, R.D., Ericsson, L.H., Neurath, H., Walsh, K.A. & Titani, K. (1981) Proc. Natl. Acad. Sci. USA 78, 5381-5385.

Bokranz, M. & Klein, A. (1987) Nucleic Acids Res. 15, 4350-4351.

Bowen, D., Littlechild, J.A., Fothergill, J.E., Watson, H.C. & Hall, L. (1988) Biochem. J. 254, 509-517.

Bright, J.R. (1990) personal communication.

Bright, J.R., Byrom, D., Hough, D.W., Towner, P. & Danson, M.J. (1990) J. Gen. Microbiol. (in press).

Brown, J.W., Daniels, C.J. & Reeve, J.N. (1989) CRC Crit. Rev. Microbiol. 16, 287-338.

Christiansen, C., Freundt, E.A. & Black, F.T. (1975) Int. J. Syst. Bacteriol. 25, 99-101.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) Science 209, 1406-1414.

Crabb, J.W., Murdock, A.L., Suzuki, T., Hamilton, J.W., McLinden, J.H. & Amelunxen, R.E. (1981) J. Bacteriol. 145, 503-512.

Cubellis, M.V., Rozzo, C., Montecucchi, P. & Rossi, M. (1990) Gene (in press).

Cubellis, M.V., Rozzo, C., Nitti, G., Arnone, M.I., Marino, G. & Sannia, G. (1989) Eur. J. Biochem. 186, 375-381.

Danson, M.J. (1988) Adv. Microbiol. Physiol. 29, 165-231.

Danson, M.J., Black, S.C., Woodland, D.L. & Wood, P.A. (1985) FEBS Lett. 179, 120-124.

Darland, G., Brock, T.D., Samsonoff, W. & Conti, S.F. (1970) Science 170, 1416-1418.

DeLange, R.J., Green, G.R. & Searcy, D.G. (1981a) J. Biol. Chem. 256, 900-904.

DeLange, R.J., Williams, L.C. & Searcy, D.G. (1981b) J. Biol.Chem. 256, 905-911.

Donald, L.J. & Duckworth, H.W. (1987) Biochem. Cell Biol. 65, 930-938.

Donald, L.J., Molgat, G.F. & Duckworth, H.W. (1989) J. Bacteriol. 171, 5542-5550.

Else, A.J., Danson, M.J. & Weitzman, P.D.J. (1988) Biochem. J. 254, 437-442.

Evans, C.T., Owens, D., Casazza, J.P. & Srere, P.A. (1989) Biochem. Biophys. Res. Commun. 164, 1437-1445.

Evans, C.T., Owens, D.D., Sumegi, B., Kispal, G. & Srere, P.A. (1988) Biochemistry 27, 4680-4686.

Fabry, S., Lang, J., Niermann, T., Vingron, M. & Hensel, R. (1989) Eur. J. Biochem. 179, 405-413.

Fabry, S., Lehmacher, A., Bode, W. & Hensel, R. (1988) FEBS Lett. 237, 213-217.

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablan, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. & Woese, C.R. (1980) Science 209, 457-463.

Grossebüter, W. & Görisch, H. (1985) System. Appl. Microbiol. 6, 119-124.

Grunstein, M. & Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.

Grütter, M.G., Hawkes, R.B. & Mathews, B.W. (1979) Nature 277, 667-669.

Handford, P.A., Ner, S.S., Bloxham, D.P. & Wilton, D.C. (1988) Biochim. Biophys. Acta 953, 232-240.

Hathaway, J.A. & Atkinson, D.E. (1965) Biochem. Biophys. Res. Commun. 20, 661-665.

Henneke, C.M., Danson, M.J., Hough, D.W. & Osguthorpe, D.J. (1989) Prot. Engng. 2, 597-604.

Hensel, R., Zwickl, P., Fabry, S., Lang, J. & Palm, P. (1989) Can. J. Microbiol. 35, 81-85.

Hocking, J.D. & Harris, J.I. (1976) in Zuber, H. (ed.):"Enzymes and Proteins from Thermophilic Microorganisms", 121-133. Birkhäuser Verlag, Basel & Stuttgart.

Honka, E., Fabry, S., Niermann, T., Palm, P. & Hensel, R. (1990) Eur. J. Biochem. 188, 623-632.

Hough, D.W. & Danson, M.J. (1989) Lett. Appl. Microbiol 9, 33-39.

Huber, R., Kurr, M., Jannasch, H.W. & Stetter, K.O. (1989) Nature 342, 833-834.

Huber, R, Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B. & Stetter, K.O. (1986) Arch. Microbiol. 144, 324-333.

Ikai, A. (1980) J. Biochem. 88, 1895-1898.

Jones, W.J., Nagle, D.P.,jr. & Whitman, W.B. (1987) Microbiol. Rev. 51, 136-177.

Kandler, O. (ed) (1982): "Archaebacteria". Gustav Fischer Verlag (Stuttgart).

Karpusas, M., Branchaud, B. & Remington, S.J. (1990) Biochemistry 29, 2213-2219.

Klenk, H.-P., Haas, B., Schwass, V., & Zillig, W. (1986) J. Mol. Evol. 24, 167-173

Krebs, H.A. & Johnson, W.A. (1937) Enzymologia 4, 148-156.

Krebs, H.A., Gurin, S. & Eggleston, L.V. (1952) Biochem. J. 51, 614-628.

Kushner, D.J. (1985) in Woese, C.R. & Wolfe, R.S. (eds.): "The Bacteria" 8, 171-214. Acad. Press Inc. (London) Ltd.

Lake, J.A. (1987) Mol. Biol. Evol. 4, 167-191.

Lake, J.A. (1988) Nature 331, 184-186.

Lake, J.A. (1989) Can. J. Microbiol. 35, 109-118.

Lane, D.J., Pace, B. Olsen, G.J., Stahl, D.A., Sogin, M.L. & Pace, N.R. (1985) Proc. Natl. Acad. Sci. USA 82, 6955-6959.

Langworthy, T.A., Holzer, G., Zeikus, J.G. & Tornabene, T.G. (1983) Syst. Appl. Microbiol. 4, 1-17.

Langworthy, T.A., Smith, P.F. & Mayberry, W.R. (1972) J. Bacteriol. 112, 1193-1200.

Lechner, K. & Bock, A. (1987) Mol. Gen. Genet. 208, 523-528.

Lowenstein, J.M. (1967) in Greenberg, D.M. (ed.): "Metabolic Pathways" 1, 146-270 (third edition). Acad. Press Inc. (London) Ltd.

Lynen, F. & Reichert, E. (1951) Angew. Chem. 63, 47-48.

Margulis, L. (1970): "Origin of eukaryotic cells". Yale University Press (New Haven, Conn.).

Ner, S.S., Bhayana, V., Bell, A.W., Giles, I.G., Duckworth, H.W. & Bloxham, D.P. (1983) Biochemistry 22, 5243-5249.

Neumann, H., Gierl, A., Tu, J., Leibrock, J., Staiger, D. & Zillig, W. (1983) Mol. Gen. Genet. 192, 66-72.

Olsen, G.J. & Woese, C.R. (1989) Can. J. Microbiol. 35, 119-123.

Oyaizu, H., Debrunner-Vossbrinck, B., Mandelco, L., Studier, J.A. & Woese, C.R. (1987) System. Appl. Microbiol. 9, 47-53.

Pauptit, R.A., Karlsson, R., Picot, D., Jenkins, J.A., Niklaus-Reimer, A.-S. & Jansonius, J.N. (1988) J. Mol. Biol. 199, 525-537.

Perutz, M.F. (1978) Science 201, 1187-1191.

Perutz, M.F. & Raidt, H. (1975) Nature 255, 256-259.

Pond, J.L., Langworthy, T.A. & Holzer, G. (1986) Science 231, 1134-1136.

Prangishvilli, D., Zillig, W., Gierl, A., Biesert, L. & Holz, I. (1982) Eur. J. Biochem. 122, 471-477.

Pribnow, D. (1975) J. Mol. Biol. 99, 419-443.

Ree, H.K. & Zimmerman, R.A. (1990) Nucleic Acids Res. (in press).

Ree, H.K., Cao, K., Thurlow, D.L. & Zimmermann, R.A. (1989) Can. J. Microbiol. 35, 124-133.

Remington, S., Wiegand, G. & Huber, R. (1982) J. Mol. Biol. 158, 111-152.

Roberts, R.B., Cowie, D.B., Britten, R., Bolten, E. & Abelson, P.H. (1953) Proc. Natl. Acad. Sci. USA 39, 1013.

Rosenkrantz, M., Alam, T., Kim K.-S., Clark, B.J., Srere, P.A. & Guarente, L.P. (1986) Mol. Cell. Biol. 6, 4509-4515.

Rossi, M. (1988), personal communication.

Sanangelantoni, A.M., Barbarini, D., Di Pasquale, G., Cammarano, P. & Tiboni, O. (1990) Mol. Gen. Genet. 221, 187-194.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Searcy, D.G. (1975) Biochim. Biophys. Acta 395, 535-547.

Searcy, D.G. (1986) System. Appl. Microbiol. 7, 198-201.

Searcy, D.G. & Doyle, E.K. (1975) Int. J. Syst. Bacteriol. 25, 286-289.

Searcy, D.G. & Whatley, F.R. (1982) in Kandler, O. (ed): "Archaebacteria", 245-257. Gustav Fischer Verlag (Stuttgart).

Segerer, A., Langworthy, T.A. & Stetter, K.O. (1988) System. Appl. Microbiol. 10, 161-171.

Shine, J. & Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342-1346.

Smith, L.D. (1989) PhD thesis, University of Bath.

Smith, L.D., Stevenson, K.J., Hough, D.W. & Danson, M.J. (1987) FEBS Lett. 225, 277-281

Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.

Srere, P.A. (1972) Curr. Top. Cell. Regul. 5, 229-283.

Srere, P.A., Brazil, H. & Gonen, L. (1963) Acta Chem. Scand. 17, S129-S134.

Stein, D.B. & Searcy, D.G. (1978) Science 202, 219-221.

Stern, J.R., Shapiro, B. & Ochoa, S. (1950) Nature 166, 403-404.

Stetter, K.O. & Zillig, W. (1985) in Woese, C.R. & Wolfe, R.S. (eds.): "The Bacteria" 8, 85-170. Acad. Press Inc. (London) Ltd.

Stetter, K.O., König, H. & Stackebrandt, E. (1983) System. Appl. Microbiol. 4, 535-551.

Suissa, M., Suda, K. & Schatz, G. (1984) EMBO J. 3, 1773-1781.

Tautz, D. & Renz, M. (1983) Anal. Biochem. 132, 14-19.

Taylor, G.L. (1990) in Crabbe, M.J.C. (ed): "Enzyme Biotechnology", 61-96. Ellis Horwood Ltd. (Chichester, Sussex)

Thomm, M. & Wich, G. (1988) Nucleic Acids Res. 16, 151-163.

Tiboni, O., Sanangelantoni, A.M., Cammarano, P., Cimino, L., Di Pasquale, G. & Sora, S. (1989) System. Appl. Microbiol. 12, 127-133.

Tu, J. & Zillig, W. (1982) Nucleic Acids Res. 10, 7231-7245.

Unger, E.A., Hand, J.M., Cashmore, A.R. & Vasconcelos, A.C. (1989) Plant Mol. Biol. 13, 411-418.

Wakabayashi, S., Fujimoto, N., Wada, K., Matsubara, H., Kerscher, L. & Oesterhelt, D. (1983) FEBS lett. 162, 21-24.

Walker, J.E., Wonacott, A.J. & Harris, J.I. (1980) Eur. J. Biochem. 108, 581-586.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner J., Hirose, T. & Itakura, K. (1979) Nucleic Acids Res. 6, 3543-3557

Weitzman, P.D.J. (1966) Biochim. Biophys. Acta 128, 213-215

Weitzman, P.D.J. (1981) Adv. Microbiol. Physiol. 22, 185-244.

Weitzman, P.D.J. & Danson, M.J. (1976) Curr. Top. Cell. Regul. 10, 161-204.

Weitzman, P.D.J. & Dunmore, P. (1969) FEBS Lett. 3, 265-267.

Weitzman, P.D.J. & Hewson, J.K. (1973) FEBS Lett. 36, 227-231.

Weitzman, P.D.J. & Jones, D. (1968) Nature 219, 270-272.

Wiegand, G. & Remington, S.J. (1986) Ann. Rev. Biophys. Biophys. Chem. 15, 97-117.

Wiegand, G., Remington, S., Deisenhofer, J. & Huber, R. (1984) J. Mol. Biol. 174, 205-219.

Williams, S.A., Slatko, B.E., Moran, L.S. & DeSimone, S.M. (1986) BioTechniques 4, 138-147.

Winnacker, E.-L. (1987): "From Genes to Clones". VCH Verlagsgesellschaft (Weinheim, West Germany).

Woese, C.R. (1981) Scientific American 244, 94-106.

Woese, C.R. (1985) in Schleifer, K.H. & Stackebrandt, E., (ed.): "Evolution of Prokaryotes" FEMS symposium 29, 1-30. Academic Press Inc. (London) Ltd.

Woese, C.R. (1987) Microbiol. Rev. 51, 221-271.

Woese, C.R. & Fox, G.E. (1977) Proc. Natl. Acad. Sci. USA 74, 5088-5090.

Woese, C.R. & Olsen, G.J. (1986) System. Appl. Microbiol. 7, 161-177.

Woese, C.R., Kandler, O. & Wheelis, M.L. (1990) Proc. Natl. Acad. Sci. USA 87, 4576-4579.

Woese, C.R., Magrum, L.J. & Fox, G.E. (1978) J. Mol. Evol. 11, 245-252.

Wood, D.O., Williamson, L.R., Winkler, H.H. & Krause, D.C. (1987) J. Bacteriol. 169, 3564-3572.

Wright, J.A., Maeba, P. & Sanwal, B.D. (1967) Biochem. Biophys. Res. Commun. 29, 34-38.

Yang, D., Kaine, B.P. & Woese, C.R. (1985) System. Appl. Microbiol. 6, 251-256.

Zillig, W., Klenk, H.-P., Palm, P., Leffers, H., Pühler, G., Gropp, F. & Garret, R.A. (1989) Endocytobiosis & Cell Res. 6, 1-25.

Zillig, W., Palm, P., Reiter, W.-D., Gropp, F., Pühler, G. & Klenk, H.-P. (1988) Eur. J. Biochem. 173, 473-482.